
ИЗДАТЕЛЬСТВО
МИР

General relativity

An Einstein centenary survey
edited by

S. W. HAWKING
Professor of Gravitational Physics
University of Cambridge

W. ISRAEL
Professor of Physics
University of Alberta

CAMBRIDGE UNIVERSITY PRESS
CAMBRIDGE
LONDON NEW YORK MELBOURNE

ОБЩАЯ ТЕОРИЯ ОТНОСИТЕЛЬНОСТИ

Под редакцией
С. Хокинга и В. Израэля

Перевод с английского
под редакцией
д-ра физ.-мат. наук Я. А. Смородинского и
д-ра физ.-мат. наук В. Б. Брагинского

Москва «Мир»
1983

ББК 22.313
О-28
УДК 530.12:531.51

Общая теория относительности: Пер. с англ./Под ред.
О-28 С. Хокинга, В. Израэля.— М.: Мир, 1983.— 455 с., ил.

Коллективная монография, написанная группой известных физиков-теоретиков СССР, США и Великобритании. Обсуждаются теоретические и экспериментальные основы общей теории относительности, альтернативные релятивистские гравитационные теории, проблемы современной астрофизики и космологии, в том числе физическая природа квазаров, пульсаров и черных дыр.

Для научных работников, аспирантов и студентов старших курсов.

О $\frac{1704020000-367}{041(01)-83}$ 46—83, ч. 1

ББК 22.313
530.1

Редакция литературы по физике

ПРЕДИСЛОВИЕ РЕДАКТОРА ПЕРЕВОДА

Статьи, переводы которых включены в настоящий сборник, взяты из большого тома, выпущенного Кембриджским университетом к 100-летию со дня рождения Эйнштейна. Нам пришлось для сокращения объема пожертвовать рядом статей и из 16 статей английского издания отобрать лишь восемь.

Один из редакторов английского издания, С. Хокинг получил известность как автор идеи об испарении черной дыры — рождении пар в сильных гравитационных полях вблизи горизонта. Вместе с Р. Пенроузом, автором одной из статей сборника, он внес в развитие общей теории относительности новые геометрические идеи, оказавшие большое влияние на современный этап развития теории.

Статьи сборника дают хорошее представление об основных направлениях исследований, как они представлялись в 1979 г. Эта область исследуется весьма интенсивно, и за прошедшие годы многие результаты получили дальнейшее развитие. По-видимому, наибольшие успехи достигнуты в теории суперсимметрии и супергравитации — теории, пытающейся объединить различные взаимодействия микромира и включить в такую, пока еще не совершенную модель, гравитацию. Этому направлению, совсем не представленному в настоящей книге, посвящен сборник «Геометрические идеи в физике», выходящий в издательстве «Мир» [1]. Кроме того, полезно обратить внимание читателя на книгу Н. Биррелла и П. Девиса «Квантованные поля в искривленном пространстве-времени», перевод которой также готовится к печати в издательстве «Мир» [2], и книгу К. Уилла «Теория и эксперимент в гравитационной физике» [3] с подробной дискуссией вопросов, частично затронутых в статье этого автора, включенной в настоящий сборник.

К юбилею А. Эйнштейна было издано много сборников [4—6], перевод одного из них, «Альберт Эйнштейн и теория гравитации», выпущен издательством «Мир» в 1979 г. Юбилей великого ученого совпал с периодом, когда теория относительности, считавшаяся почти законченной наукой, испытала новое рождение. Два направления — релятивистская космология и астрофизика и геометрия гравитационных полей и их особенности — вызывают особый интерес. Хорошим введением в эту область и должны служить статьи предлагаемого сборника.

Перевод выполнили: канд. физ.-мат. наук. А. Г. Полнарев (предисловие, гл. I), канд. физ.-мат. наук Э. А. Тагиров (гл. II, V—VII) и канд. физ.-мат. наук И. Г. Дымникова (гл. III, IV, VIII).

Я. А. Смородинский

ЛИТЕРАТУРА

1. Геометрические идеи в физике. Сб. статей 1978—1980 гг. Пер. с англ.— М:. Мир, 1983.
2. *Birrell N. D., Davies P. C. W.*, Quantum fields in curved space, Cambridge, 1982.
3. *Will C.*, Theory and experiment in gravitational physics, Cambridge, 1981.
4. Astrofisica e cosmologia gravitaziona. Quante e Relativita, Firenza, Italia, 1979 (готовится к изданию сокращенный перевод на русский язык).
5. Relativity, Quanta and Cosmology, vols. I—II, Johnson Reprint Corporation, New York—San Francisco, London, 1979.
6. General Relativity and Gravitation, ed. by A. Held, vols. I—II, Plenum Press, New York—London. 1980.

ПРЕДИСЛОВИЕ

Физику тех лет 1879 год показался бы ничем не примечательным. Он мог бы прочесть скорее со скептицизмом, чем с интересом, статью Иозефа Стефана в Sitzungsberichte der Akademie der Wissenschaften in Wien (Протоколы Венской академии наук), автор которой, исходя из довольно туманных экспериментальных фактов, заключил, что энергия, излучаемая абсолютно поглощающим телом, пропорциональна четвертой степени его температуры. В физику незаметно вошла новая фундаментальная постоянная. Однако чтобы это осознать, потребовались еще 20 лет и проницательность Планка.

Вероятно, наш физик с бо́льшим интересом воспринял бы в том же году появление восьмого тома девятого издания Encyclopaedia Britannica, содержащего статью Джеймса Клерка Максвелла «Эфир». Но он не мог бы тогда знать о письме Максвелла Д. П. Тодду, директору Nautical Almanac, в котором Максвелл спрашивал, не знает ли он о каких-нибудь аномалиях в моментах затмений спутников Юпитера. «Дело в том,— писал Максвелл,— что они, насколько мне известно, дают нам единственный способ определить направление и величину скорости движения Солнца относительно светоносной среды..., а при использовании земных методов определения скорости света свет возвращается обратно по тому же пути, так что движение Земли относительно эфира изменяет время двойного прохождения на величину, зависящую от отношения квадрата скорости Земли к квадрату скорости света; однако эта величина слишком мала, чтобы ее можно было заметить». Письмо Максвелла было опубликовано в Nature вскоре после его смерти в ноябре 1879 г. и попалось на глаза молодому преподавателю Военно-морской академии США Альберту А. Майкельсону. Это письмо с пометкой «Кавендишская лаборатория, Кембридж, 19 марта 1879 г.», воспроизводится в конце нашего предисловия. Пятью днями ранее, 14 марта 1879 г., в г. Ульме родился Альберт Эйнштейн. Так два родника, которые впоследствии должны были дать жизнь квантовой теории и теории относительности, появились на свет одновременно с человеком, который в критические годы направит потоки от этих родников по правильному руслу.

С самого ее появления в 1916 г. общая теория относительности считалась чрезвычайно трудной теорией. Ее развитие за последние 20 лет было столь феноменальным, что в книге, подобной этой,

невозможно избежать довольно пространного изложения технических деталей. Однако главное впечатление, которое неизбежно возникает у каждого,— это чувство изумления перед той волшебной, неизъяснимо рациональной Вселенной, которая открылась нам гением одного человека. Никто не может без трепета вспомнить свое первое столкновение с этим миром, как будто созданным фантазией Льиюса Кэрролла, где пространство искривлено, время служит четвертым измерением и честнейшие свидетели лукаво противоречат друг другу в ответ на самые элементарные вопросы о том, что происходит, где и когда. Возможно, самый бесценный дар, который Альберт Эйнштейн оставил в наследство нашему столетию,— это то, что он вернул миру биллиардных шаров его ореол глубины и таинственности.

Несмотря на то что со дня рождения Эйнштейна прошло более ста лет, сейчас самое время вновь обозреть его величайшее достижение — общую теорию относительности. За последние 20 лет достигнут огромный прогресс как в понимании теории, так и в ее экспериментальном подтверждении, в результате чего общая теория относительности из интеллектуальной достопримечательности превратилась в повседневный рабочий инструмент для астрофизиков, и ее эффекты следует учитывать даже в столь практических делах, как конструирование спутниковых систем навигации. Сейчас мы имеем довольно полную картину теории, по крайней мере до того момента, когда она потребует модификации, чтобы включить в себя квантовые гравитационные эффекты. Большинство статей данного сборника описывает различные стороны этой классической, т. е. не квантовой, теории. Однако классическая теория также предсказывает, что должны существовать пространственно-временные сингулярности, т. е. области, где кривизна пространства-времени становится неограниченно большой. Можно было бы ожидать, что в таких ситуациях доминирующую роль будут играть квантовые гравитационные эффекты. В настоящее время этот вопрос интенсивно исследуется, и ему посвящено несколько последних статей нашего сборника. Конечной целью является осуществление мечты Эйнштейна о создании полной и согласованной теории, которая объединила бы все законы физики.

Составляя этот сборник, мы стремились обеспечить обзор современного состояния исследований по общей теории относительности, который был бы доступен для неспециалистов. Изложение некоторых вопросов, таких, как проблема начальных данных, с неизбежностью оказалось усложненным и наполненным множеством уравнений и доказательств. Однако неспециалист не встретит трудностей при чтении других статей, посвященных экспериментальным проверкам, поискам гравитационного излучения, астрофизике черных дыр и т. п.

Редакторы благодарны авторам, написавшим эти статьи за короткий срок, секретарям Джуди Фелла и Мэри Ю, Яну Бейли и

Ричарду Мэбли за помощь в составлении предметного указателя, а также редакторам издательства «Кембридж юниверсити пресс», которые проделали огромную работу по подготовке рукописи к печати.

Август 1978 г.

С. Хокинг
В. Израэль

«О возможных способах определения движения Солнечной системы в светоносном эфире». Заметка покойного проф. Дж. К. Максвелла, содержащаяся в письме к м-ру Д. П. Тодду, директору Nautical Almanac, Вашингтон, США. Представлено проф. Стоксом, секретарем Королевского общества.

М-р Тодд был столь добр, что передал мне копию прилагаемого письма и любезно разрешил использовать его по моему усмотрению.

Поскольку заметка в Encyclopaedia Britannica, на которую ссылается Максвелл, была очень краткой — ограничивалась одной фразой, а обсуждаемый вопрос представляет огромный интерес, я счел целесообразным передать письмо в Королевское общество.

Из проведенных м-ром Хиггинсом исследований компоненты относительной скорости вдоль луча зрения нашего Солнца и некоторых звезд можно заключить о допустимом эффекте, не превышающем половины секунды. Несомненно, что его обнаружение составляет очень деликатную задачу. Однако из всего, что нам известно *a priori*, следует, что это движение должно быть значительно более заметным, поэтому предложенная идея важна сама по себе вне зависимости от возможности реальных измерений.

В своем письме ко мне м-р Тодд замечает: «Я рассматриваю это сообщение как чрезвычайно важное, хотя (как Вы заметите, если имеете возможность ознакомиться с моим ответом), вероятно, пройдет немало времени, прежде чем нам удастся получить таблицы движения спутников Юпитера, достаточно точные, чтобы их можно было использовать для практических проверок».

Я не считал целесообразной задержку публикации письма в ожидании, что какие-нибудь материалы, касающиеся этого вопроса, могут быть найдены среди бумаг Максвелла.

(Копия)

Кавендишская лаборатория
Кембридж
19 марта 1879 г.

Сэр,

Я с большим удовольствием получил таблицы движения спутников Юпитера, которые Вы столь любезно переслали мне. Ваш интерес к системе Юпитера дает мне смелость задать вопрос, занимались ли Вы какими-либо специальными исследованиями кажущегося запаздывания моментов затмений, связанного с геоцентрическим положением Юпитера.

Мне говорили, что наблюдения такого рода несколько вышли из моды из-за других методов определения скорости света, но они, насколько мне известно, дают нам единственный способ определить направление и величину скорости движения Солнца относительно светоносной среды. Даже если бы мы были уверены в теории аберрации, мы смогли бы лишь получить разности в положении звезд, а при использовании земных методов определения скорости света свет возвращается обратно по тому же пути, так что движение Земли относительно эфира изменяет время двойного прохождения на величину, зависящую от отношения квадрата скорости Земли к квадрату скорости света; однако эта величина слишком мала, чтобы ее можно было заметить.

Но если JE есть расстояние от Юпитера до Земли, l — геоцентрическая долгота, а l' — долгота и λ — широта направления, по которому Солнце движется

через эфир со скоростью v, и если V — скорость света, а t — время его прохождения от J к E, то

$$JE = [V - v\cos\lambda\cos(l - l')]\,t.$$

Сравнивая значение t, когда Юпитер находится в разных знаках Зодиака, можно определить l' и $v\cos\lambda$.

Я не знаю, как определить λ, кроме как по планете с орбитой, имеющей большое наклонение к эклиптике. Следует заметить, что в то время как определение скорости света V с помощью этого метода зависит от изменений расстояния JE, т. е. от диаметра земной орбиты, определение $v\cos\lambda$ зависит от самой величины JE, которая гораздо больше.

Однако ни один метод не может быть использован без хороших таблиц движения спутников, и поскольку я не астроном, я не знаю, предпринимались ли по аналогии с наблюдениями при помощи таблиц Дамуазье какие-либо попытки рассмотреть член $v\cos\lambda$.

Я позволил себе написать Вам, поскольку обсуждаемый вопрос выходит за рамки исследований, проведенных теми, кто не занимался специально спутниками Юпитера.

В статье «Эфир», опубликованной в девятом издании Encyclopaedia Britannica, я собрал все известные мне факты об относительном движении эфира и тел, движущихся в нем, и показал, что никаких заключений об этом относительном движении нельзя сделать на основании всех наблюдаемых в настоящее время явлений, кроме затмений спутников планет, чем далее расположенных, тем лучше.

Если Вам известна какая-либо работа в этом направлении, Ваша лично или других авторов, я счел бы за честь получить об этом сообщение.

Искренне Ваш

(подпись) *Дж. К. Максвелл*

Д. П. Тодд, эсквайр

I. ТЕОРИЯ ГРАВИТАЦИИ И ЭКСПЕРИМЕНТ [1]

К. М. Уилл [2, 3]

1. ВВЕДЕНИЕ

Уже более полувека общая теория относительности — поистине монумент в память Альберта Эйнштейна. Она навсегда изменила наши представления о пространстве и времени, заставив нас вплотную подойти к проблемам рождения и конечной судьбы Вселенной. И все же, несмотря на свое огромное влияние на научную мысль, общая теория относительности первоначально опиралась на крайне скудные эмпирические данные. Почти сорок пять лет справедливость этой теории подтверждалась совокупностью экспериментов, которые по современным критериям в лучшем случае можно считать лишь качественными. Три «классических теста» составили эмпирический базис общей теории относительности: дополнительное смещение перигелия Меркурия, согласующееся с предсказанием теории, но допускающее, по крайней мере частично, объяснение другими причинами; отклонение светового луча Солнцем, для которого измерения давали значения в интервале от половины до удвоенной величины, предсказываемой теорией; и, наконец, гравитационное красное смещение, наблюдавшееся в спектральных линиях белых карликов с той же низкой точностью, что и отклонение света Солнцем, и которое, более того, едва ли можно вообще рассматривать как истинную проверку теории относительности.

Однако благодаря революции в астрономии и технике, относящейся к 60-м — 70-м годам нашего столетия, сопоставление теории относительности с экспериментом достигло беспрецедентного уровня точности. Такие технические достижения, как создание атомных часов, измерения расстояний до планет и искусственных спутников Земли с помощью радиолокаторов и лазеров, радиоинтерферометрия со сверхдлинными базами, высокочувствительные датчики смещений — вот неполный список арсенала средств, благодаря которым систематические проверки теории гравитации с высокой точностью стали делом почти привычным. Измерения ничтожно малых эффектов, предсказываемых общей теорией относительности для Солнеч-

---

[1]) Работа частично финансирована Национальным управлением по аэронавтике и исследованию космического пространства (NSG7204S1) и Национальным научным фондом (RHY76-21454).

[2]) *C. M. Will*, Department of Physics, Stanford University, Stanford California, USA.

[3]) Стипендиат фонда Альфреда П. Слоана.

ной системы, с точностями, составляющими доли процента, стали реальностью.

Стремление к осуществлению столь точных проверок гравитационных теорий усилилось в последние два десятилетия, когда были открыты экзотические астрономические объекты, в которых релятивистские гравитационные эффекты играют, по-видимому, первостепенную роль. В результате открытия таких объектов, как квазары, пульсары, компактные рентгеновские источники и космическое микроволновое фоновое излучение, для теоретиков-астрофизиков стало необходимым иметь полную уверенность в справедливости той единственной теории гравитации, которую они используют в построении моделей указанных выше явлений. Эта необходимость стала еще более настоятельной, когда был предложен ряд конкурирующих теорий гравитации, главным образом теория Бранса — Дикки, которые претендуют на то, чтобы их рассматривали как жизнеспособные альтернативы общей теории относительности. На протяжении 60-х и 70-х годов активность и степень изощренности при построении альтернативных теорий гравитации были столь велики, что какое-то время общая теория относительности казалась всего лишь одной из множества ничем не уступающих друг другу теорий. Дело дошло до того, что в течение 60-х и начале 70-х годов была разработана «теория гравитационных теорий», предназначенная для предельно объективного исследования и классификации всех теорий гравитации. Начатая Робертом Г. Дикки и Кеннетом Нордтведтом, эта «теория теорий» могла быть очень полезной также и при анализе новых очень точных измерений и при планировании экспериментов в будущем, возможных благодаря дальнейшему техническому прогрессу.

К сегодняшнему дню общая теория относительности выдержала все экспериментальные проверки, которым была подвергнута, а многие ее конкуренты остались за бортом, однако программа сопоставления теории гравитации с экспериментом еще далека от завершения. Точность тестов в Солнечной системе будет непрерывно возрастать, причем наибольшее продвижение планируется на 80-е годы. Однако не обязательно ограничивать проверки теории гравитации рамками Солнечной системы. Возможно, вскоре удастся проводить эксперименты в совершенно новых областях явлений, затронув при этом такие аспекты теории гравитации, которые недоступны обычным тестам в Солнечной системе. Речь идет о гравитационно-волновом эксперименте, о тестах по звездным системам (двойной пульсар), о космических тестах и лабораторных экспериментах. Можно только строить догадки, будет ли впредь общая теория относительности выживать при каждом новом ее сопоставлении с экспериментом, но так или иначе благодаря общей теории относительности мы уже достигли той глубины понимания природы гравитации и эффектов, связанных с ней, что уже это само по себе говорит о непреходящей ценности предвидения Альберта Эйнштейна.

Мы начнем обзор (разд. 2) с анализа основ теории гравитации — принципов эквивалентности, фундаментальных критериев жизнеспособности гравитационной теории и экспериментов, лежащих в основе этих критериев. Один из центральных выводов разд. 2 состоит в том, что правильная, жизнеспособная теория гравитации, по всей вероятности, должна быть «метрической» теорией. В разд. 3 мы обращаемся к проверкам в Солнечной системе, используя «теорию теорий», известную под названием параметризованного постньютоновского (ППН) формализма, который охватывает большинство метрических теорий гравитации и наилучшим образом приспособлен к анализу явлений в Солнечной системе. В разд. 4 обсуждается гравитационное излучение как возможное средство проверки гравитационной теории. Двойной пульсар — новое поле деятельности по проверке теории с помощью «звездных систем» — исследуется в разд. 5. В разд. 6 описаны космологические тесты гравитационной теории, в разд. 7 даны выводы.

В этом обзоре мы придерживаемся единиц и условных обозначений, принятых в книге Мизнера, Торна и Уилера [1] (в дальнейших ссылках сокращенно МТУ). Некоторые части обзора основаны на лекциях автора, прочитанных в 1972 г. в Варенне, «Теоретические методы экспериментальной гравитации» [2] (в дальнейших ссылках сокращенно ТМЭГ). Читателя, который захочет подробнее ознакомиться с математическими методами и вычислительными приемами, упомянутыми в этом обзоре, мы отсылаем к ТМЭГ [2] и МТУ [1], гл. 38—40. Имеется ряд других полезных работ [3—9]. В качестве полного обзора ситуации в эксперименте на 1961 г. см. работу Бертотти и др. [10].

2. ПРИНЦИПЫ ЭКВИВАЛЕНТНОСТИ И ОСНОВЫ ГРАВИТАЦИОННОЙ ТЕОРИИ

Принцип эквивалентности сыграл важную роль в развитии гравитационной теории. Исаак Ньютон считал этот принцип краеугольным камнем механики и даже посвятил первый параграф своих «Начал» детальному обсуждению этого принципа. Кроме того, он упомянул об экспериментах с маятником, проведенных им для проверки этого принципа. Эйнштейн положил этот принцип в основу общей теории относительности. Но только сравнительно недавно мы достигли более глубокого понимания значения принципа эквивалентности для гравитационной теории и эксперимента. Главным образом благодаря работе Роберта Г. Дикки мы стали понимать, что принцип (или лучше сказать принципы) эквивалентности и эксперименты типа эксперимента Этвеша, эксперименты по измерению красного смещения и другие затрагивают более фундаментальные аспекты гравитационной теории, чем собственно общая теория относительности. Эта точка зрения является частью системы, которая впоследствии получила название системы Дикки (п. 2.1); она

позволяет рассматривать на самом фундаментальном уровне природу пространства-времени и гравитации. В рамках этой системы ставятся следующие вопросы: все ли тела под действием гравитации приобретают одинаковые ускорения? Является ли пространство по своим внутренним свойствам локально изотропным? Какие типы полей, если они есть, связаны с гравитацией — скалярные поля, векторные поля, тензорные поля . . .? На основе этой системы был получен ряд фундаментальных критериев, которым должна удовлетворять любая потенциально жизнеспособная гравитационная теория (п. 2.2), предложен и путь анализа ключевых экспериментов, которые образуют эмпирический фундамент этих критериев (п. 2.3). Эти критерии привели к важным для гравитационной теории следствиям, согласно предположению, основанному на работе покойного Леонарда Шиффа. Догадка Шиффа (п. 2.4) состоит в утверждении, что любая теория гравитации, удовлетворяющая фундаментальным критериям жизнеспособности, должна обязательно быть теорией метрической. Догадка Шиффа и система Дикки породили целый ряд конкретных теоретических формализмов (таких, как система $TH\varepsilon\mu$, п. 2.5) для сравнения и противопоставления метрических и неметрических теорий гравитации, для обработки экспериментов, проверяющих критерии жизнеспособности, и для доказательства догадки Шиффа (п. 2.5).

2.1. СИСТЕМА ДИККИ

Система Дикки для анализа экспериментальных проверок гравитации изложена в приложении 4 к «лезушским лекциям» Дикки [8]. В рамках этой системы делаются два основных предположения о природе гравитации.

1. Пространство-время — это четырехмерное многообразие, каждая точка которого соответствует физическому событию. *Apriori* не обязательно, чтобы это многообразие обладало какой-либо метрикой или аффинной связностью.

2. Уравнения, описывающие гравитацию, и математические величины, входящие в них, должны выражаться в виде, не зависящем от конкретного выбора используемых координат, т. е. должны иметь ковариантную форму.

Система Дикки особенно полезна при проектировании и интерпретации экспериментов, в которых ставится вопрос, какие типы полей связаны с гравитацией [8, 11]; ТМЭГ [2], § 2; МТУ [1], (§ 38.7). Из физики элементарных частиц мы твердо знаем, что заведомо существует по крайней мере одно поле, описываемое симметричным тензором второго ранга, которое сводится к метрике Минковского η, когда гравитационными эффектами можно пренебречь. Эксперимент Хьюза — Дривера исключает существование более чем одного поля, описываемого тензором второго ранга, непосредственно взаимодействующего с веществом, и различные эксперимен-

ты по обнаружению дрейфа относительно эфира исключают возможность существования векторного поля, непосредственно взаимодействующего с веществом. (Однако эти эксперименты не исключают существование векторных и тензорных полей, связанных только с гравитацией или с собственной гравитацией вещества [12].) До сих пор ни один эксперимент не смог исключить или обнаружить существование скалярного поля, хотя целый ряд экспериментов позволил наложить ограничения на конкретные скалярно-тензорные теории (см. разд. 3). Однако польза от системы Дикки этим не ограничивается.

2.2. КРИТЕРИИ ЖИЗНЕСПОСОБНОСТИ ГРАВИТАЦИОННОЙ ТЕОРИИ

Широкий и объективный подход, воплощенный в системе Дикки, позволил теоретикам-релятивистам сформулировать ряд фундаментальных критериев, которым должна удовлетворять любая гравитационная теория, если она претендует на жизнеспособность. Два из этих критериев являются чисто теоретическими, тогда как в основе других лежит эксперимент.

1. Теория должна быть полной, т. е. она должна быть способна объяснить из «первых принципов» результат любого представляющего интерес эксперимента. Для теории недостаточно *постулировать*, что тела из различных материалов падают с одним и тем же ускорением. Теория должна включать в себя и объединять в единое целое полный набор законов электромагнетизма и квантовой механики, которые могут быть использованы при расчете детального поведения тел в гравитационных полях. Это требование полноты, конечно, не распространяется на тот круг явлений, которые связаны с квантовой гравитацией, с теорией сильных и слабых взаимодействий, с сингулярностями и т. д., т. е. на тот круг явлений, для которых даже специальную или общую теорию относительности нельзя считать до конца разработанной или полной.

2. Теория должна быть самосогласованной. Гравитационная теория считается самосогласованной, если ее предсказание относительно результата любого эксперимента является однозначным, т. е. если предсказания, полученные двумя различными методами вычислений, всегда дают одни и те же результаты, например искривление луча света можно рассчитать либо в рамках геометрической оптики, полученной предельным переходом из уравнений Максвелла, либо воспользовавшись уравнениями для движения пробных частиц в пределе нулевой массы покоя.

3. Теория должна быть релятивистской, т. е. в пределе, когда гравитация «выключается», становясь несущественной по сравнению с другими физическими взаимодействиями, негравитационные законы физики должны сводиться к законам специальной теории относительности.

4. Теория должна давать правильный ньютоновский предел, т. е. в пределе слабых гравитационных полей и медленных движений она должна воспроизводить ньютоновские законы (с нужной степенью точности), которые проявляют себя, например, в экспериментах Кавендиша, в движениях планет и в строении звезд.

В ТМЭГ [2] (§ 3.2) и в работе [13] приведены неполные списки теорий, в которых нарушается один или несколько из этих четырех критериев.

5. Теория должна включать в себя слабый принцип эквивалентности (СПЭ), который гласит: *если незаряженное пробное тело поместить в некотором начальном событии в пространстве-времени и сообщить ему там некоторую начальную скорость, то последующая мировая линия этого тела не зависит от его внутренней структуры и химического состава* (см. работу [14], где этот принцип обсуждается и формулируется более подробно). Этот принцип иногда формулируют как принцип универсальности свободного падения (УСП) (МТУ [1], § 38.3).

6. Теория должна включать в себя принцип универсальности гравитационного красного смещения, который гласит: *гравитационное красное смещение между парой идентичных идеально правильных часов в двух событиях пространства-времени не зависит от их структуры и химического состава* [15].

Прежде чем обсуждать теоретические следствия из этих критериев, обратимся к их экспериментальному обоснованию.

2.3. ЭКСПЕРИМЕНТЫ, ЗАТРАГИВАЮЩИЕ ОСНОВЫ ГРАВИТАЦИИ

Целый ряд различных экспериментов как лабораторных, так и в Солнечной системе составляет эмпирический фундамент для критериев 3—6.

а. Проверки специальной теории относительности

Множество лабораторных экспериментов при высоких энергиях дало возможность проверить и перепроверить справедливость специальной теории относительности в пределе, когда гравитационными эффектами можно пренебречь. Диапазон таких экспериментов широк: от прямых проверок замедления времени до тонких эффектов, предсказываемых лоренц-инвариантными квантовыми теориями поля (см., например, МТУ [1], § 36.4, [11, 16, 17]).

Эксперимент Хьюза — Дривера, или эксперимент по «изотропии инертной массы» [18, 19], можно рассматривать как проверку специальной теории относительности (а также тест на существование второго космического тензорного поля [20, 21]). Этот эксперимент наложил ограничение, составляющее $2 \cdot 10^{-16}$ эВ, на относительный сдвиг подуровней $m={}^3/_2$ и $m={}^1/_2$ (расщепление во внешнем магнит-

ном поле) ядерного энергетического уровня $1p_{3/2}$ в ^{7}Li. Полная энергия связи этих подуровней порядка 1 МэВ, так что экспериментальное ограничение составляет несколько единиц на 10^{-22}. Если бы имело место нарушение лоренц-инвариантности, например, в электродинамике, то за счет движения Земли относительно особой универсальной системы покоя, выделяемой нарушенной лоренц-инвариантностью, обязательно наблюдалось бы указанное расщепление. Если через δ_L обозначить безразмерный параметр, соответствующий «отклонению» от лоренц-инвариантности (конкретный вид δ_L см. в п. 2.5), то полученное из эксперимента Хьюза-Дривера ограничение на δ_L принимает вид [22]

$$|\delta_L| < 10^{-13}\,(w/30\ \text{км/с})^{-2}, \qquad (1)$$

где w — скорость Земли относительно универсальной системы отсчета.

б. Ньютоновский предел

Огромная совокупность эмпирических данных подтверждает справедливость ньютоновской гравитационной теории (НГТ) по крайней мере в качестве приближения к «истинной» релятивистской теории гравитации. Наблюдения за движением планет и космических кораблей согласуются с НГТ вплоть до уровня относительной точности порядка 10^{-8}, при которой уже могут быть обнаружены постньютоновские эффекты. Лабораторный эксперимент Кавендиша дает более слабое подтверждение НГТ для малых расстояний между гравитирующими телами. Однако совсем недавно НГТ подверглась суровым испытаниям. Было высказано предположение, что может существовать массивная короткодействующая компонента гравитационного взаимодействия, приводящая к модификациям НГТ, которые можно описать гравитационной «постоянной» $G(r)$, зависящей от расстояния между гравитирующими телами [23—26]. Недавно Лонг [27] сообщил об экспериментальных результатах, указывающих на вариации в $G(r)$ на расстояниях от 5 до 30 см, имеющие величину

$$\varepsilon \equiv G^{-1} r\, dG/dr \approx (20 \pm 5) \cdot 10^{-4},$$

где ошибка составляет одно стандартное отклонение. Готовятся попытки провести аналогичные эксперименты в других лабораториях [28, 29]; одновременно с этим Миккельсон и Ньюмен [30] провели детальное исследование ограничений, налагаемых на ε, для расстояний, превышающих 100 см, на основе разнообразных астрономических наблюдений. Необходимо провести в будущем дополнительные эксперименты, чтобы укрепить нашу уверенность в справедливости ньютоновского предела.

в. Проверки слабого принципа эквивалентности

Непосредственная проверка СПЭ состоит в сравнении ускорения объектов лабораторных масштабов, имеющих различный химический состав. Было проведено множество очень точных экспериментов по проверке СПЭ, начиная с экспериментов с маятниками Ньютона, Бесселя и Поттера и кончая классическими экспериментами с крутильными маятниками Этвеша, Дикки и Брагинского и сотр. (подробное обсуждение экспериментальных проблем см. в обзоре [8]). Эксперименты по проверке СПЭ для отдельных атомов и элементарных частиц пока неубедительны или, лучше сказать, не очень точны, за исключением эксперимента для нейтронов [31] (относительно ситуации для электронов см. работу [32]). В табл. 1 указаны различные

Таблица 1

Проверки слабого принципа эквивалентности

Эксперимент	Ссылка	Метод	Проверяемые вещества	Ограничение на $\|\eta\|$
Ньютона	Ньютон, 1686 [33]	Маятник	Различные	10^{-3}
Бесселя	Бессель, 1832 [34]	Маятник	Различные	$2\cdot10^{-5}$
Этвеша	Этвеш, Пекар, Фекете, 1922 [35]	Крутильный маятник	Различные	$5\cdot10^{-9}$
Поттера	Поттер, 1923 [36]	Маятник	Различные	$2\cdot10^{-5}$
Реннера	Реннер, 1935 [37]	Крутильный маятник	Различные	$2\cdot10^{-9}$
Принстонский	Ролл, Кротков, Дикки, 1964 [38]	Крутильный маятник	Алюминий и золото	10^{-11}
Московский	Брагинский, Панов, 1971 [39]	Крутильный маятник	Алюминий и платина	10^{-12}
Кёстера	Кёстер, 1976 [31]	Свободное падение	Нейтроны	$3\cdot10^{-4}$
Стэнфордский	Уорден, 1976 [40]	Магнитный подвес	Ниобий, Земля	$2\cdot10^{-5}$
Орбитальный [a]	Уорден, 1976 [40]	Свободное падение на орбите	Различные	От 10^{-15} до 10^{-18}

[a] Эксперимент, который только предстоит провести.

эксперименты и приведены ограничения, которые они накладывают на разницу в ускорении a между объектами с различным химическим составом (A и B), определяемую как

$$\eta \equiv (a_A - a_B)/(^1/_2)(a_A + a_B).$$

Будущие усовершенствованные проверки СПЭ должны уменьшить шумы, обусловленные тепловыми и сейсмическими эффектами, поэтому их, возможно, следует проводить в космосе, используя криогенную технику. Ожидаемые ограничения на величину η в таких экспериментах лежат в интервале от 10^{-15} до 10^{-18} [40, 41].

г. Проверки принципа УГКС

Измерения гравитационного красного смещения служат проверками принципа УГКС. В типичном эксперименте смещение частоты или длины волны $z=\Delta\nu/\nu=-\Delta\lambda/\lambda$ между одинаковыми атомными стандартами частоты, которые расположены в точках с различным гравитационным потенциалом, измеряется в зависимости от разницы потенциалов ΔU. По многим соображениям удобно считать, что эксперименты измеряют параметр α, определяемый соотношением

$$z=(1+\alpha)\,\Delta U. \tag{2}$$

Если справедлив принцип УГКС, то α не зависит от природы изучаемых часов; более того, имеется ряд теоретических аргументов (догадка Шиффа, см. пп. 2.4, 2.5) в пользу того, что при справедливости УГК параметр α должен равняться нулю. В табл. 2 собраны краткие сведения о важнейших экспериментах по красному смеще-

Таблица 2

Проверки универсальности гравитационного красного смещения

Эксперимент	Ссылка	Метод	Ограничение на $\|\alpha\|$
Паунда—Ребки—Снайдера	Паунд, Ребка, 1960 [42], Паунд, Снайдер, 1965 [43]	Падение фотонов, испущенных мессбауэровскими излучателями	10^{-2}
Браулта	Браулт, 1962 [44]	Солнечные спектральные линии	$5\cdot10^{-2}$
Дженкинса	Дженкинс, 1969 [45]	Кварцевые часы на борту спутника ГЕОС-1	$9\cdot10^{-2}$
Снайдера	Снайдер, 1972 [46]	Солнечные спектральные линии	$6\cdot10^{-2}$
По запаздыванию часов на реактивном самолете А	Хафеле, Китинг, 1972 [47, 48]	Часы на цезиевых пучках на борту реактивного самолета	10^{-1}
По запаздыванию часов на реактивном самолете Б	Эли и др., 1977 [49]	Рубидиевые часы на борту реактивного самолета	$2\cdot10^{-2}$
НАСА—САО ракетный эксперимент по красному смещению	Вессот, Левин, 1976 [50]	Водородный мазер на ракете	$2\cdot10^{-4}$
Нулевой эксперимент по красному смещению	Турнюр, Уилл, 1975 [51] Уилл, 1977 [52]	Сравнение водородного мазера с АССР	10^{-2}
Зонд, проходящий вблизи Солнца [a]	Нордтведт, 1977 [53]	Водородный мазер или АССР на спутнике	10^{-6}

[a] Эксперименты, которые только предстоит провести.

нию, проведенных с 1960 г. (обзор экспериментов до 1960 г. см. в работе [10]), и упомянуто несколько экспериментов, которые, возможно, будут проведены в ближайшие годы. Первой и наиболее известной очень точной проверкой явился эксперимент Паунда — Ребки — Снайдера, в котором измерялось смещение частоты гамма-фотонов, испущенных ^{57}Fe, при их движении от основания к вершине гарвардской башни и наоборот. Высокая точность достигалась благодаря использованию эффекта Мессбауэра для получения узкой резонансной линии. В других экспериментах измерялось смещение спектральных линий от Солнца в гравитационном поле Солнца. Некоторые проверки включали атомные часы, поднимаемые на высоту самолетами, ракетами и спутниками.

Однако в последнее время наступила новая эра для экспериментов по красному смещению; это связано с развитием стандартов частоты сверхвысокой стабильности: порядка 10^{-15}—10^{-16} при временах усреднения от 10 до 100 с и более. В качестве примера можно привести водородные лазерные часы [54], часы, представляющие собой автогенератор, стабилизированный сверхпроводящим резонатором (АССР) [55, 56], и охлажденные криогенными методами монокристаллы из диэлектриков, таких, как кремний и сапфир [57]. Впервые такой эксперимент был проведен Национальным управлением по аэронавтике и исследованию космического пространства совместно со Смитсонианской астрофизической обсерваторией (НАСА — САО) в июне 1976 г., это был ракетный эксперимент по изучению красного смещения. Водородные мазерные часы были запущены на ракете на высоту около 10 000 км и их частота сравнивалась с частотой таких же часов на Земле. В этом же эксперименте использовалась высокая стабильность частоты водородных мазерных часов (порядка 10^{-15} при временах усреднения в пределах 100 с) и проводилось слежение за смещением частоты в зависимости от высоты. Специальная схема обработки данных [54] позволяет полностью исключить все эффекты, связанные с линейным доплеровским смещением вследствие движения ракеты; одновременно получаемые данные использовались для определения положения и скорости последней ступени ракеты (с тем чтобы рассчитать разность потенциалов ΔU и квадратичное доплеровское смещение). Грубый предварительный анализ данных позволил получить α в пределах $2 \cdot 10^{-4}$ [50]. (По странному совпадению ракета, которой воспользовались для подъема мазера, на стартовой площадке имела высоту 22,6 м, т. е. почти точно высоту гарвардской башни.) В межпланетном варианте подобного эксперимента предполагается выведение стабильных часов (H-мазер или АССР) на орбиту вокруг Солнца с очень большим эксцентриситетом (с минимальным расстоянием до Солнца порядка четырех солнечных радиусов); в таком эксперименте величину α можно проверить с точностью 10^{-6} [53], что позволило бы искать эффекты красного смещения «второго порядка» [58].

Успехи в создании стабильных часов сделали возможным про-

ведение нового типа эксперимента по измерению красного смещения, который мог бы служить непосредственной проверкой принципа УГКС: это «нулевой» эксперимент по красному смещению, в котором сравниваются часы двух различных типов, помещенные рядом в одной и той же лаборатории. В одном из таких экспериментов [51, 52] предполагается сравнивать водородные мазерные часы с АССР и следить за дневными вариациями в их относительных частотах по мере того, как за счет вращения Земли лаборатория попадает в области то более высокого, то более низкого гравитационного потенциала Солнца. Такие вариации имеют следующий вид:

$$\nu_H/\nu_{\text{АССР}} = A\,[1 - 3\cdot 10^{-13}\,(\alpha_H - \alpha_{\text{АССР}})\cos(2\pi t/1 \text{ солнечный день})],$$

где A — постоянное среднее отношение между частотами двух часов [52]. Зависящий от химического состава коэффициент $\alpha_H - \alpha_{\text{АССР}}$ может быть ограничен величиной порядка 10^{-2}.

2.4. ПРИНЦИП ЭКВИВАЛЕНТНОСТИ И ПРЕДПОЛОЖЕНИЕ ШИФФА

В п. 2.2 мы описали два принципа эквивалентности, СПЭ и УГК, которые должны входить в любую потенциально жизнеспособную теорию гравитации. Третий принцип, эйнштейновский принцип эквивалентности (ЭПЭ), также играет важную роль в понимании основ гравитационной теории. Этот принцип гласит: 1) *справедливы СПЭ и принцип УГКС* и 2) *результат любого негравитационного эксперимента, проведенного в локальной, свободно падающей системе отсчета, не зависит от того, где и когда во Вселенной этот эксперимент проводился, и не зависит от скорости системы отсчета* (обсуждение этого принципа см. в работе [14]). От ЭПЭ всего один шаг до постулатов *метрической теории гравитации:* 1) *пространство-время обладает метрикой* $g_{\mu\nu}$; 2) *мировые линии пробных тел являются геодезическими этой метрики*; 3) *удовлетворяется ЭПЭ*; *при этом все негравитационные законы в любой свободно падающей системе отсчета сводятся к законам специальной теории относительности.* Таким образом, ЭПЭ указывает на принципиальное различие между метрическими теориями гравитации и неметрическими теориями, в которых не выполнен один или несколько метрических постулатов.

Предположение Шиффа состоит в утверждении, что *любая теория гравитации, допускаемая фундаментальными критериями жизнеспособности* (п. 2.2), *с необходимостью удовлетворяет ЭПЭ.* В таком виде предположение Шиффа является видоизменением для классического случая ранее высказанного им в 1960 г. квантовомеханического предположения [59]. Его интерес к этому предположению вновь вспыхнул в ноябре 1970 г. в ожесточенном споре с Кипом Торном на конференции по экспериментальной гравитации, проходившей в Калифорнийском технологическом институте. К несчастью, его

преждевременная кончина в январе 1971 г. прервала возобновленную им деятельность в этом направлении.

Если предположение Шиффа правильно, то эксперименты Этвеша и эксперименты по измерению гравитационного красного смещения можно считать непосредственным эмпирическим фундаментом для ЭПЭ и, следовательно, для трактовки гравитации как геометрического явления, связанного с искривленностью пространства-времени. Вот почему в последнее время так много усилий направлено на «доказательство» предположения Шиффа в рамках особых математических систем, которые охватывают как метрические, так и неметрические теории гравитации, оставаясь при этом достаточно простыми, чтобы позволить проводить конкретные вычисления.

2.5. МЕТРИЧЕСКИЕ И НЕМЕТРИЧЕСКИЕ ТЕОРИИ ГРАВИТАЦИИ; СИСТЕМА $TH\varepsilon\mu$

Первая успешная попытка доказательства предположения Шиффа была предпринята Лайтманом и Ли [60]. Они разработали систему, названную $TH\varepsilon\mu$-формализмом, которая охватывает все метрические теории гравитации и многие неметрические теории (табл. 3). В рамках этой системы удается рассмотреть поведение заряженных частиц (речь идет только об электромагнитных взаимодействиях) во внешнем статическом сферически-симметричном (ССС) гравитационном поле, которое характеризуется потенциалом U. Движение заряженных частиц описывается двумя произвольными функциями потенциала $T(U)$ и $H(U)$, а поведение электромагнитного поля во внешнем потенциале (уравнения Максвелла, видоизмененные с учетом гравитации) — двумя функциями потенциала $\varepsilon(U)$ и $\mu(U)$. Вид функций T, H, ε и μ в разных теориях различен, но в любой метрической теории для всех U выполняется следующее соотношение:

$$\varepsilon=\mu=(H/T)^{1/2}. \qquad (3)$$

И наоборот, в любой теории из класса теорий, удовлетворяющих соотношению (3), уравнения электромагнитного поля можно привести к «метрическому» виду. Затем Лайтман и Ли [60] рассчитали в явном виде ускорение падения «пробного» тела, состоящего из взаимодействующих заряженных частиц, и обнаружили, что ускорение падения не зависит от структуры тела (СПЭ) в том и только том случае, если выполняется соотношение (3). Другими словами, СПЭ $\Rightarrow$ ЭПЭ (т. е. из СПЭ следует ЭПЭ), и предположение Шиффа тем самым проверено, по крайней мере в рамках ограничений, присущих данному формализму (по поводу обратного утверждения ЭПЭ $\Rightarrow$ СПЭ см. работу [61]). Уилл [15] из $TH\varepsilon\mu$-формализма получил модифицированное с учетом гравитации уравнение Дирака, определил гравитационное красное смещение, испытываемое различными атомными часами, и установил, что красное смещение не зави-

Таблица 3

$TH\varepsilon\mu$-формализм [60]

А. *Лагранжиан для частиц с массой m_{0a} и зарядом e_a:*

$$L=\sum_a\int\left[-m_{0a}\left(T-Hv_a^2\right)^{1/2}+e_aA_\mu\left(x_a^\nu\right)v_a^\mu\right]dt+$$
$$+(8\pi)^{-1}\int\left[\varepsilon\mathbf{E}^2+\mu^{-1}\mathbf{B}^2\right]d^3x\,dt$$

Б. *Гравитационное поле:* $U(\mathbf{x})=M/r$, где M — постоянная, r — расстояние до источника в декартовых координатах

В. *Произвольные функции:* $T(U)$, $H(U)$, $\varepsilon(U)$, $\mu(U)$; ЭПЭ удовлетворяется в том и только том случае, если $\varepsilon=\mu=(H/T)^{1/2}$ для любых U; вдали от источника поля функции принимают значения T_∞, H_∞, ε_∞, μ_∞, определяемые современными космологическими граничными условиями; отнормировав координаты и заряды так, чтобы в настоящее время $T_\infty=H_\infty=\varepsilon_\infty=1$, из эксперимента Хьюза — Дривера имеем ограничение на μ_∞, даваемое неравенством $|\varepsilon_\infty\mu_\infty T_\infty/H_\infty-1|<10^{-13}$

Г. *Разложение в случае слабого поля и неметрические параметры* (в предположении $T_\infty=H_\infty=\varepsilon_\infty=\mu_\infty=1$):

$$T=1-2U+2\beta U^2+\ldots,$$
$$H=1+2\gamma U+{}^3/_2\delta\cdot U^2+\ldots,$$
$$\varepsilon=1+(1+\gamma-\Gamma_0)\,U+\frac{1}{2}\left(3+2\gamma-2\beta-\gamma^2+\frac{3}{2}\delta-\Gamma_1\right)U^2+\ldots,$$
$$\mu=1+(1+\gamma-\Lambda_0)\,U+\frac{1}{2}\left(3+2\gamma-2\beta-\gamma^2+\frac{3}{2}\delta-\Lambda_1\right)U^2+\ldots$$

Параметры γ, β, δ равны аналогичным ППН-параметрам (п. 3.2); (Γ_0, Λ_0), (Γ_1, Λ_1) ... являются мерой отклонения от ЭПЭ в каждом порядке по U.

сит от природы часов (УГКС) в том и только том случае, если выполняется соотношение (3); другими словами, УГКС $\Rightarrow$ ЭПЭ, и тем самым проверен другой аспект предположения Шиффа.

Сущность $TH\varepsilon\mu$-формализма заключается в «законе привязки к метрике» (3), который получил такое название потому, что он налагает ограничения на характер «привязки» электромагнитных законов и уравнений движения частиц к гравитационному полю при условии, что эти законы должны быть «метрическими». Экспериментальные проверки принципа эквивалентности можно рассматривать поэтому как проверки закона привязки к метрике, по крайней мере для тех случаев, когда речь идет об электромагнитных эффектах. Поскольку такие эксперименты проводятся в Солнечной системе, где гравитация слаба ($U<10^{-6}$), наибольший интерес представляет версия закона привязки в слабом поле. Воспользовавшись произвольными параметрами (табл. 3), можно разложить функции T, H, ε и μ в ряд по степеням U. Пары параметров Γ_0 и Λ_0, Γ_1 и Λ_1 и т. д. служат мерой отклонения от закона привязки к метрике

(3) в каждом порядке по U. В любой метрической теории гравитации $\Gamma_i=\Lambda_i=0$, тогда как в любой неметрической теории хотя бы один параметр Γ_i или Λ_i не равен нулю.

В пределе слабого поля ускорение падения неточечного сферического пробного тела имеет вид [60, 62]

$$\mathbf{a}=(m_p/m)\nabla U,$$
$$m_p/m=1-\Gamma_0(E_e/m)-\Lambda_0(E_m/m)+$$
$$+[\text{члены, включающие } \Gamma_1,\ \Lambda_1]+\ldots, \quad (4)$$

где m_p и m — пассивная гравитационная и инертная массы пробного тела, E_e и E_m — собственная электростатическая и собственная магнитостатическая энергии тела (более подробно см. работу [62]). Эксперименты Этвеша налагают ограничения на члены в (4), приводящие к нарушению СПЭ, что в конце концов приводит к следующим ограничениям на неметрические параметры Γ_0, Λ_0, Γ_1, Λ_1 и т. д.:

$$|\Gamma_0|<4\cdot 10^{-10}, \quad |\Lambda_0|<6\cdot 10^{-6},$$
$$|\Gamma_1|<4\cdot 10^{-2}, \quad |\Lambda_1|<600.$$

Эти ограничения являются достаточно сильными, чтобы исключить целый ряд неметрических теорий гравитации, которые до этого считались жизнеспособными; сюда входят теории Белинфанте и Свихарта, Найды и Капеллы [60].

Гравитационные красные смещения, которые испытывают фотоны, испущенные различными типами источников, имеют вид (2) [15, 52]; при этом

$$\alpha=\begin{cases}-3\Gamma_0+\Lambda_0 & \text{(сверхтонкий переход в водороде, Н-мазерные часы),}\\ -\frac{1}{2}(3\Gamma_0+\Lambda_0) & \text{(электромагнитная мода в резонаторе, АССР-часы),}\\ -2\Gamma_0 & \text{(фононная мода в твердом теле, высокодобротный диэлектрический кристалл; основной переход в водороде).}\end{cases}$$

Таким образом, ракетный эксперимент по измерению красного смещения, проведенный НАСА и САО, налагает ограничение на комбинацию параметров $|3\Gamma_0-\Lambda_0|<2\cdot 10^{-4}$; нулевой эксперимент по красному смещению, в котором сравниваются водородные мазерные часы и АССР-часы (п. 2.3, г), наложил бы ограничение на комбинацию параметров $|\alpha_H-\alpha_{АССР}|=|(^3/_2\ (\Gamma_0-\Lambda_0)|$. Эксперимент Хьюза—Дривера (п. 2.3, а) также можно проанализировать в рамках $TH\varepsilon\mu$-формализма. Лагранжиан в $TH\varepsilon\mu$-формализме, приведенный в табл. 3, является лоренц-инвариантным при условии $\varepsilon\mu T/H=1$; нарушение этого условия приводит к такой анизотропии инерт-

ной массы, которая оказалась несовместимой с экспериментом. Детальные расчеты [22] приводят к следующему пределу для величины δ_L [см. (1)]:

$$\delta_L = | \varepsilon\mu T/H - 1 | < 10^{-13},$$

где принято $w = 30$ км·с$^{-1}$.

В $TH\varepsilon\mu$-формализме заложено много ограничений: рассматриваются только сферически-симметричные статические (ССС) поля, только электромагнитные взаимодействия и специальный вид лагранжиана. Было предпринято несколько попыток ослабить эти ограничения. Ни [63] обобщил формализм, выйдя за рамки ССС-полей, но, продолжая работать с лагранжианом взаимодействия электромагнитного поля и частиц весьма специального, хотя и более общего вида, чем в $TH\varepsilon\mu$-формализме. Он убедился, что предположение Шиффа справедливо, пока нет никакого псевдоскалярного гравитационного поля, взаимодействующего с максвелловским полем. В присутствии некоторого неметрического псевдоскалярного поля для пробных тел по-прежнему справедлив СПЭ, но тела испытывают аномальные крутящие моменты, связанные с моментами в распределении их внутренней электромагнитной энергии. $TH\varepsilon\mu$-формализм является частным случаем уравнений Ни.

Нордтведт [64] воспользовался мысленным экспериментом, с тем чтобы, исходя только из закона сохранения энергии и локальной лоренц-инвариантности, продемонстрировать тесную связь между СПЭ и УГКС. В этом мысленном эксперименте система, находившаяся в квантовом состоянии A, переходит в состояние B, испуская квант с частотой ν. Квант падает с высоты H во внешнем гравитационном поле, его частота испытывает смещение, принимая некоторое значение ν', в то время как система в состоянии B падает с ускорением g_B. Сообщив системе в состоянии B квант с частотой ν' и кинетическую энергию $m_B g_B H$, приобретенную системой во время падения, восстанавливаем систему в состоянии A. Оставшейся энергии $m_A g_A H$ должно в точности хватить на то, чтобы поднять систему, находящуюся в состоянии A, в первоначальное положение. (Предположение о локальной лоренц-инвариантности допускает отождествление инертных масс m_A и m_B с полной энергией каждого из тел.) Если g_A g_B зависят от той доли внутренней энергии состояний, которая участвовала в переходе из состояния A в состояние B, так что

$$g_A = g(1 + \alpha E_A/m_A), \quad g_B = g(1 + \alpha E_B/m_B), \quad E_A - E_B = h\nu$$

(нарушение СПЭ), то закон сохранения энергии требует соответствующего нарушения УГКС, которое (в наименьшем порядке по $h\nu/m$) должно иметь вид

$$z = (\nu' - \nu)/\nu = (1 + \alpha)\, gH.$$

В рамках $TH\varepsilon\mu$-формализма это соотношение было проверено

Хогеном [65] для кулоновской электростатической внутренней энергии и для внутренней энергии тонкой и сверхтонкой структуры при условии выполнения лоренц-инвариантности, т. е. при $\varepsilon\mu T/H = 1$. Если потребовать выполнения СПЭ ($\alpha=0$), то мы возвращаемся к классическому «выводу» обычного гравитационного красного смещения, которым пользовались Дикки [8] и другие.

Попытки ослабить ограниченность $TH\varepsilon\mu$-формализма, связанную с рассмотрением только электромагнитного поля, наталкиваются на неопределенности и методические сложности теорий слабого и сильного взаимодействий. Однако Хоген и Уилл [66] показали, что если слабые взаимодействия вызывают нарушения СПЭ вида

$$m_P/m = 1 + \Gamma_w (E_w/m),$$

где E_w — вклад слабых взаимодействий в энергию ядра, Γ_w — параметр, являющийся мерой нарушения СПЭ, то московский вариант эксперимента Этвеша налагает предел $|\Gamma_w| < 5\cdot 10^{-3}$; другими словами, слабые взаимодействия подчиняются СПЭ с точностью, лучшей чем 1%. Тем самым были опровергнуты более ранние утверждения, что эксперименты такого типа не могут служить проверкой гравитационных эффектов, связанных со слабыми взаимодействиями [8, 40, 41, 67, 68].

Одно из важнейших условий справедливости ЭПЭ, которое, однако, лежит за пределами современного $TH\varepsilon\mu$-формализма,— это постоянство фундаментальных негравитационных постоянных на интервалах времени космологического масштаба (обсуждение гравитационной «постоянной» мы отложим до п. 6.1). Такое постоянство служит непосредственной проверкой той части ЭПЭ (п. 2.4), в которой говорится, что «результат любого негравитационного эксперимента . . . не зависит от того, *когда* этот эксперимент проводился». Мы не будем давать обзор различных теорий и предложений, возникших после того, как Дирак допустил переменность фундаментальных констант (подробный обзор и библиографию см. в [69], вместо этого перечислим данные самых последних наблюдений (табл. 4). Эти наблюдения охватывают широкий круг явлений: от сравнения спектральных линий в далеких галактиках и квазарах [70] до измерения распространенности изотопов элементов в Солнечной системе [69] и сравнения лабораторных атомных часов [55]. Недавно Шлайактер [71] существенно усилил ограничения на вариации электромагнитной, слабой и сильной констант взаимодействия, изучая продукты деления в «естественном реакторе» в Окло (Габон, Африка), в котором, по всей видимости, около 2 млрд. лет назад протекали самоподдерживающиеся реакции деления. Сравнивая ядерные сечения, полученные по распространенности изотопов, возникших в ходе этих реакций, с сечениями, измеряемыми в наши дни, он получил ограничения, указанные в последнем столбце табл. 4.

Таблица 4

Ограничения на космологическое изменение негравитационных постоянных

Постоянная k	Ограничение на $\dot{k}/k$ в единицах обратного хаббловского времени $2\cdot10^{10}$ лет [H_0= 55 км/(с·Мпс)]	Метод	Литература	Ограничение по реактору в Окло, Шлайактер, 1976 [71]
Размерные константы:				
$\hbar c$	$2\cdot10^{-2}$	Зависимость энергии от длины волны для старых и молодых фотонов	Солхейм, Барнес, Смит, 1976 [72], Баум, Флорентин-Нильсен, 1976 [73]	
Безразмерные константы:				
постоянная тонкой структуры: $\alpha = e^2/\hbar c$	$4\cdot10^{-4}$	β-распад ^{187}Re за геологическое время	Дайсон 1972 [69]	10^{-7}
	$8\cdot10^{-2}$	Тонкая структура MgII и линия 21 см в радиоисточнике при $z=0,5$	Вольфе, Браун, Робертс, 1976 [70]	
	$8\cdot10^{-2}$	Сравнение АССР-часов и часов на цезиевых пучках	Турнюр, Стейн, 1976 [74]	
Константа слабого взаимодействия: $\beta = g_f m^2 c/\hbar^3$	2	Скорости распада ^{187}Re, ^{40}K	Дайсон, 1972 [69]	$2\cdot10^{-4}$
Отношение массы электрона к массе протона: m_e/m_p	1	Массовое смещение в спектральных линиях квазаров ($z\sim2$)	Пагель, 1977 [75]	
Гиромагнитное отношение для протона: $g_p m_e/m_p$	10^{-1}	Mg II, линия 21 см	Вольфе, Браун, Робертс, 1976 [70]	
Сильные взаимодействия: g_s^2	$8\cdot10^{-2}$	Стабильность ядер	Дэвис 1972 [76]	$8\cdot10^{-9}$

2.6. ВЫВОДЫ

Убедительное экспериментальное подтверждение фундаментальных критериев жизнеспособности гравитационных теорий в совокупности с теоретическими аргументами, такими, как предположение Шиффа, укрепляют нашу уверенность в том, что какой бы ни была истинная теория гравитации, она должна быть метрической теорией. В оставшейся части настоящего обзора мы будем считать, что это предположение правильно. Однако следует иметь в виду, что это положение опирается на фундамент, который может оказаться весьма непрочным, и непрекращающиеся экспериментальные и теоретические усилия крайне необходимы для поддержания надежности основ гравитационной теории.

3. ПОСТНЬЮТОНОВСКАЯ ГРАВИТАЦИЯ В СОЛНЕЧНОЙ СИСТЕМЕ

Вникая в смысл постулатов метрических теорий гравитации (п. 2.4), мы обнаруживаем важнейшую особенность: какой бы сложной ни была теория, с какими бы дополнительными гравитационными или космологическими полями она ни имела дела, вещество и негравитационные поля реагируют только на метрику $g_{\mu\nu}$. Роль всех других полей, которые включает в себя данная теория, сводится лишь к участию в создании кривизны пространства-времени, связанной с этой метрикой. Вещество может порождать эти поля, и эти поля вместе с веществом могут генерировать метрику, но непосредственного обратного воздействия на вещество эти поля оказывать не могут. Вещество реагирует только на метрику. С этой точки зрения метрика $g_{\mu\nu}$ является первичным теоретическим понятием, и все различие между разными метрическими теориями сводится лишь к тому, как конкретно в данной теории вещество генерирует метрику.

Сравнение метрических теорий гравитации друг с другом и с экспериментом становится особенно простым, если перейти к постньютоновскому пределу слабых полей и медленных движений. Оказывается, в этом предельном случае все теории отличаются друг от друга лишь предсказываемыми значениями десяти безразмерных параметров, так называемых ППН-параметров (п. 3.1). Используя получающийся в этом предельном случае параметризованный постньютоновский (ППН) формализм, можно изучать и классифицировать конкурирующие метрические теории гравитации (п. 3.2). Поскольку в Солнечной системе гравитация слаба, ППН-формализм идеально удобен для анализа экспериментов в Солнечной системе, в которых могут быть обнаружены постньютоновские эффекты. Каждый эксперимент, следовательно, можно рассматривать как измерение соответствующего ППН-параметра или комбинации параметров и тем самым как проверку предсказаний конкурирующих теорий. К важнейшим экспериментам относятся эксперименты по откло-

нению света и запаздыванию времени (п. 3.3), проверки эффекта Нордтведта с помощью лазерной локации Луны (п. 3.4), измерения смещения перигелия (п. 3.5), геофизические и планетарные проверки эффектов, связанных с выделенной системой отсчета или выделенным положением в пространстве (п. 3.6), эксперименты с гироскопами и другие эксперименты, связанные с прецессией (п. 3.7), а также лабораторные эксперименты (п. 3.8).

3.1. ПАРАМЕТРИЗОВАННЫЙ ПОСТНЬЮТОНОВСКИЙ (ППН) ФОРМАЛИЗМ

В пределе слабого поля и медленных движений пространственно-временна́я метрика $g_{\mu\nu}$, предсказываемая почти любой метрической теорией гравитации, обладает одной и той же структурой. Метрику можно записать в виде разложения «вблизи» метрики Минковского ($\eta_{\mu\nu}=\mathrm{diag}(-1, 1, 1, 1)$) по безразмерным гравитационным потенциалам различной степени малости. Эти потенциалы строятся с помощью переменных, описывающих вещество (табл. 5),

Таблица 5

Параметризованный постньютоновский формализм

А. *ППН-параметры:*

$$\gamma,\ \beta,\ \xi,\ \alpha_1,\ \alpha_2,\ \alpha_3,\ \zeta_1,\ \zeta_2,\ \zeta_3,\ \zeta_4$$

Б. *Метрика:*

$$g_{00}=-1+2U-2\beta U^2-2\xi\Phi_W-(\zeta_1-2\xi)\mathcal{A}+(2\gamma+2+\alpha_3-\zeta_1-2\xi)\Phi_1+$$
$$+2(3\gamma-2\beta+1+\zeta_2+\xi)\Phi_2+2(1+\zeta_3)\Phi_3+2(3\gamma+3\zeta_4-2\xi)\Phi_4-$$
$$-(\alpha_1-\alpha_2-\alpha_3)w^2U-\alpha_2 w^i w^j U_{ij}+(2\alpha_3-\alpha_1)w^iV_i,$$

$$g_{0i}=-\frac{1}{2}(4\gamma+3+\alpha_1-\alpha_2+\zeta_1-2\xi)V_i-\frac{1}{2}(1+\alpha_2-\zeta_1+2\xi)W_i-$$
$$-\frac{1}{2}(\alpha_1-2\alpha_2)w^iU-\alpha_2 w^jU_{ij},$$

$$g_{ij}=(1+2\gamma U)\delta_{ij}$$

В. *Метрические потенциалы:*
Подробные определения см. в ТМЭГ [2], § 4, и в [77].

Г. *Различия между принятым здесь вариантом и вариантом, приведенным в ТМЭГ [2]*

1. Приняты такие же, как в книге МТУ, сигнатура (—1, 1, 1, 1,) и система индексов (греческие индексы принимают значения 0, 1, 2, 3; латинские индексы принимают значения 1, 2, 3).
2. Новый символ для параметра Уайтхеда: ξ вместо ξ_W, как это было в [77].
3. Модифицированные параметры, связанные с законами сохранения и включающие теперь член, описывающий эффект Уайтхеда [78]:

$$(\zeta_1)_{\text{нов}}=(\zeta_1)_{\text{стар}}+2\xi,\quad (\zeta_2)_{\text{нов}}=(\zeta_2)_{\text{стар}}-\xi,$$
$$(\zeta_4)_{\text{нов}}=(\zeta_4)_{\text{стар}}+\frac{2}{3}\xi.$$

и ньютоновского гравитационного потенциала

$$U(\mathbf{x},\ t) = \int \rho(\mathbf{x}',\ t)\,|\,\mathbf{x}-\mathbf{x}'\,|^{-1}\,d^3\mathbf{x}'.$$

«Порядок малости» определяется в соответствии со следующими правилами: $U \sim v^2 \sim \Pi \sim p/\rho \sim O(2)$, $v^i \sim |d/dt|/|d/dx| \sim O(1)$ и т. д. Для самосогласованного описания постньютоновского предела необходимо определить g_{00} с точностью до $O(4)$, g_{0i} — с точностью до $O(3)$ и g_{ij} — с точностью до $O(2)$ (подробности см. в ТМЭГ [2]). Различные метрические теории отличаются друг от друга только численными значениями коэффициентов, стоящих перед метрическими потенциалами. В ППН-формализме вместо этих коэффициентов вводятся параметры, значения которых зависят от конкретной рассматриваемой теории. В современной версии ППН-формализма используются десять параметров, выбранных так, чтобы они являлись мерой или хотя бы отражали общие свойства метрических теорий гравитации (табл. 6). Параметры γ и β — это привычные пара-

Таблица 6

ППН-параметры и их физический смысл

Параметр (см. табл. 5)	Мерой чего он является по отношению к общей теории относительности[a]	Значение в общей теории относительности	Значение в полуконсервативных теориях	Значение в полностью консервативных теориях
γ	Как велика пространственная кривизна, создаваемая единичной покоящейся массой	1	γ	γ
β	Как велика «нелинейность» в законе суперпозиции для гравитации	1	β	β
ξ	Существуют ли эффекты, связанные с выделенным положением в пространстве	0	ξ	ξ
α_1	Существуют ли эффекты, связанные с выделенной системой отсчета	0	α_1	0
α_2		0	α_2	0
α_3		0	0	0
ζ_1	Нарушается ли закон сохранения полного импульса	0	0	0
ζ_2		0	0	0
ζ_3		0	0	0
ζ_4		0	0	0

[a] Все это справедливо лишь в стандартной ППН-калибровке и не должно рассматриваться в качестве ковариантных утверждений, иначе могут возникать недоразумения, особенно для случая параметра β: примеры см. в работах [85, 86].

метры Эддингтона [79] — Робертсона [80] — Шиффа [81], которые используются для описания «классических» проверок общей теории относительности; $\xi \neq 0$ в любой теории гравитации, предсказывающей эффекты, связанные с существованием выделенного положения в пространстве, такие, как анизотропия локальной гравитационной постоянной G_L, вызванная галактикой (такие эффекты называют также «эффектами Уайтхеда» [77, 82]); ξ соответствует ξ_W из работы [77]; параметры α_1, α_2, α_3 указывают на то, предсказывает ли теория постньютоновские эффекты, связанные с существованием выделенной системы отсчета [12, 83]; параметры α_3, ζ_1, ζ_2, ζ_3, ζ_4 указывают на то, предсказывает ли теория нарушения глобальных законов сохранения полного импульса [84]. В табл. 6 приведены значения этих параметров: 1) в общей теории относительности; 2) в любой теории гравитации, в которой справедливы законы сохранения полного импульса, такие теории называются «полуконсервативными» (любая теория, в основу которой положен принцип инвариантного действия, является полуконсервативной [78]), и 3) в любой теории, в которой, кроме того, имеется шесть глобальных законов сохранения углового момента; такие теории называют «полностью консервативными» (они автоматически предсказывают полное отсутствие постньютоновских эффектов, связанных с существованием выделенной системы отсчета). Полуконсервативные теории имеют пять свободных ППН-параметров (γ, β, ξ, α_1, α_2), тогда как в полностью консервативных теориях их всего три (γ, β, ξ). В табл. 5 приведены самые основные понятия и формулы ППН-формализма; более подробно см. об этом МТУ [1], § 39, и ТМЭГ [2], § 4.

Простейший вариант ППН-формализма был построен и исследован Эддингтоном [79], Робертсоном [80] и Шиффом [81]. Метрика в Солнечной системе принималась такой, как если бы Солнце было сферическим невращающимся телом, а планеты приближенно считались пробными телами, движущимися по геодезическим в этой метрике. Шифф [87] обобщил метрику с учетом вращения (эффект Ленза — Тирринга, п. 3.7). Но полный ППН-формализм был впервые разработан Кеннетом Нордтведтом [88], который исследовал постньютоновскую метрику системы гравитирующих точечных масс. Уилл [89] обобщил этот формализм, включив рассмотрение вещества в приближении идеальной жидкости (см. также работу Байерлайна [90]). Затем Уилл и Нордтведт [12] унифицировали ППН-формализм, и потом Уилл добавил параметр, связанный с существованием выделенного положения в пространстве, или параметр Уайтхеда. Различия между принятым здесь вариантом формализма, и вариантом Уилла — Нордтведта (приведенным в ТМЭГ [1], § 4) указаны в п. Г табл. 5. Были разработаны и другие версии ППН-формализма для описания точечных заряженных масс [91] и жидкости с анизотропными натяжениями (МТУ [1], § 39).

3.2. ЗООПАРК КОНКУРИРУЮЩИХ ТЕОРИЙ ГРАВИТАЦИИ

Одним из важных применений ППН-формализма является сравнение и классификация альтернативных метрических теорий гравитации. Список жизнеспособных теорий в последние годы время от времени менялся по мере обнаружения, главным образом благодаря ППН-подходу, новых эффектов и появления новых проверок. Примерами могут служить эффект Нордтведта [92], проверки существования выделенной системы отсчета [82, 83] и выделенного положения в пространстве [77, 82]. Эти новые проверки исключили целый ряд теорий, которые прежде считались жизнеспособными. Население зоопарка теорий время от времени меняется и потому, что предлагаются новые жизнеспособные теории.

Сегодняшний зоопарк теорий можно разбить на два широких теоретических класса. К первому классу относятся те теории гравитации, которые включают в себя только динамические гравитационные поля, тогда как ко второму классу относятся теории, в которых фигурирует первичная геометрия, т. е. в этих теориях присутствуют абсолютные или нединамические поля, такие, как плоские фоновые метрики или поля космического времени. Эти понятия подробно обсуждаются в работе [14] и кратко в МТУ [1] (§ 17.6) (см. также [93]). Следует подчеркнуть, что теории с первичной геометрией могут оставаться общековариантными, согласуясь со вторым из основных предположений системы Дикки (п. 2.1). Например, плоская фоновая метрика может быть полностью определена ковариантным образом, если потребовать, чтобы ее тензор Римана всюду обращался в нуль; координату t, соответствующую космическому времени, также можно определить ковариантным образом, задав ее как скалярное поле, градиент которого является времениподобным и ковариантно постоянным.

Теперь приведем краткое описание жизнеспособных в настоящее время метрических теорий гравитации и дадим сводку значений их ППН-параметров (табл. 7). Подробнее об этих теориях и о приемах получения их постньютоновского предела см. работу [94] и ТМЭГ [2], § 5 и приведенные там ссылки; что касается более поздних теорий, мы даем ссылки на оригинальные работы.

а. Чисто динамические теории

1. Общая теория относительности: метрика g является единственным динамическим полем, и теория не содержит никаких произвольных функций или параметров (мы не говорим здесь о космологической постоянной λ).

2. Скалярно-тензорные теории: содержат метрику g, скалярное поле ϕ и произвольную функцию связи $\omega(\phi)$, которая определяет относительную «силу» скалярного поля. В самом общем случае это теория Бергмана [95] — Вагонера [23] — Нордтведта [96] (БВН), в которой $\omega(\phi)$ — произвольная функция. Частными случаями

являются теории Бекенштейна [97] [$\omega(\phi)$ — функция специального вида от ϕ, содержащая два произвольных коэффициента r и q] и теория Бранса — Дикки [98] [$\omega(\phi)$ — константа]. Параметрами, фигурирующими в постньютоновском пределе, являются

$$\omega=\omega(\phi_0),\quad \Lambda=\left[(d\omega/d\phi)(3+2\omega)^{-2}(4+2\omega)^{-1}\right]|_{\phi_0},$$

где ϕ_0 — сегодняшняя величина ϕ на больших расстояниях от Солнечной системы, определяемая соответствующими космологическими граничными условиями. В теории Бранса — Дикки чем больше величина ω, тем менее значительны эффекты скалярного поля, и в пределе $\omega\to\infty$ теория во всех своих предсказаниях становится неотличимой от общей теории относительности. В редакции Бекенштейна БВН-теория допускает достаточно патологический вид функции $\omega(\phi)$, такой, что при сегодняшнем значении скалярного поля ϕ_0 значение ω очень велико и Λ очень мало (сегодня предсказания теории почти совпадают с предсказаниями общей теории относительности), но в прошлом или будущем, при других ϕ, ω и Λ могут принимать такие значения, что это приведет, например, к существенному видоизменению космологических моделей [97].

3. Векторно-тензорные теории [99]: содержат метрику $\mathbf{g}$ и векторное поле $\mathbf{K}$. В варианте ППН-формализма Уилла — Нордтведта [12] в постньютоновском пределе параметром является величина $\mathbf{K}_0$ вдали от Солнечной системы.

б. Теории с первичной геометрией

4. Биметрические теории содержат динамические скалярные, векторные или тензорные поля и нединамическую фоновую метрику η. В биметрической теории Розена [100—103] физическая метрика g является динамическим полем, а метрика η — риманово-плоская (недавно Розен [104] обобщил теорию на случай, когда η соответствует пространству-времени постоянной кривизны). В теории нет каких-либо произвольных параметров, однако постньютоновский предел с необходимостью должен асимптотически сливаться с космологической моделью, и тем самым определяются относительные асимптотические значения $g_{\mu\nu}$ и $\eta_{\mu\nu}$. В случае однородной изотропной космологической модели эта сшивка приводит к двум параметрам c_0 и c_1, которые даются соотношениями [105]

$$c_0+3c_1=\eta^{\mu\nu}g_{\mu\nu},\quad c_0^{-1}+3c_1^{-1}=\eta_{\mu\nu}g^{\mu\nu}, \tag{5}$$

где правые части уравнений соответствуют большим расстояниям от Солнечной системы. Тесная связь между локальной гравитацией и глобальной структурой может иметь важные следствия для жизнеспособности теорий этого типа. Например, значения ППН-параметров в этой теории и в общей теории относительности совпадают, за исключением параметра α_2, величина которого зависит от c_0 и c_1. В теории Розена существуют космологические модели, для кото-

Таблица 7

Метрические теории гравитации и значения их ППН-параметров

Теория	Произвольные функции или постоянные	Параметры, определяемые сегодняшними космологическими условиями	ППН-параметры ν	β	ξ	α_1	α_2	(α_3, ζ_i)	Можно ли сделать параметры, равными параметрам ОТО?
а. *Чисто динамические теории*									
1. Общая теория относительности (g)	Нет	Нет	1	1	0	0	0	0	
2. Скалярно-тензорные теории (g, ϕ):									
БВН	$\omega(\phi)$	ϕ_0	$\frac{1+\omega}{2+\omega}$	$1+\Lambda$	0	0	0	0	Нет[б]
Бекенштейна	$\omega(\phi)$, r, q	ϕ_0	$\frac{1+\omega}{2+\omega}$	$1+\Lambda$	0	0	0	0	Нет[б]
Бранса — Дикки	ω	ϕ_0	$\frac{1+\omega}{2+\omega}$	1	0	0	0	0	Нет[б]
3. Векторно-тензорные теории (g, K):									
Уилла — Нордтведта	Нет	K_0	1	1	0	0	$K_0^2\left(1+\frac{1}{2}K_0^2\right)^{-1}$	0	Да, но для специально выбранной космологической модели
б. *Теории с первичной геометрией*									$K_0=0$

Продолжение

Теория	Произвольные функции или постоянные	Параметры, определяемые сегодняшними космологическими условиями	ППН-параметры						Можно ли сделать параметры, равными параметрам ОТО?
			γ	β	ξ	α_1	α_2	(α_3, ζ_i)	
4. Биметрические теории:									
Розена (g, η)	Нет	c_0, c_1	1	1	0	0	$(c_0/c_1)-1$	0	Да, но для специально выбранной космологической модели ($c_0=c_1$)
Лайтмана—Ли (g, h, η)	a, f, k	ω_0, ω_1	γ' [в]	β'	0	α_1'	α_2'	0	Да, при (a, f, k, ω_0, ω_1) $=(1/4, -5/64, 1/16, 0, 0)$
Расталла (g, η, φ, K)	Нет	[а]	1	1	0	0	0	0	Да
5. Теории расслоения:									
Ни (g, η, t, φ, K)	e, $f_1(\phi)$, $f_2(\phi)$, $f_3(\phi)$, a, b, d	[а]	a	b	0	$-4\left(a+1+\frac{1}{2}e\right)$	$-(d+1)$	0	Да, при (a, b, d, e) $=(1, 1, -1, -4)$

[а] Полное соответствие между постньютоновским пределом и разумными космологическими граничными условиями еще не проведено. Указанные значения ППН-параметров соответствуют предположению, что g асимптотически равно η.

[б] См. обсуждение в п. 3.2.

[в] В теории Лайтмана—Ли γ', β', α_1' и α_2' являются сложными функциями a, f, k, ω_0 и ω_1.

рых $\alpha_2 \approx 0$ [104], но эти модели, по-видимому, противоречат данным космологических наблюдений. Мы обсудим этот вопрос подробнее в п. 6.2.

В биметрической теории Лайтмана — Ли [106] динамическим полем является тензор **h**, а физическая метрика получается затем алгебраически из тензора **h** и плоской фоновой метрики η. Теория содержит три произвольные постоянные a, f, k и два параметра, ω_0 и ω_1, определяемые из сшивки с космологией. При специальном выборе значений $(a, f, k, \omega_0, \omega_1)=\left(\frac{1}{4}, -\frac{5}{64}, \frac{1}{16}, 0, 0\right)$ ППН-параметры этой теории совпадают с ППН-параметрами общей теории относительности.

В биметрической теории Расталла [107—109] динамическими полями являются скаляр ϕ и вектор **K**, а физическая метрика получается алгебраически из ϕ, **K** и плоской фоновой метрики η. В сегодняшнем варианте эта теория не содержит каких-либо произвольных постоянных [109], однако проблема сшивки с реальной космологической моделью до сих пор не решена. Расталл [109] обнаружил, что ППН-параметры — такие же, как в общей теории относительности, если предположить, что физическая метрика асимптотически равна η.

5. Теории расслоения: включают механизм отбора выделенных пространственных сечений («слоев») во Вселенной. В самом общем случае этот механизм связан с нединамической функцией космического времени t, у которой касательный 4-вектор является времениподобным и ковариантно постоянным в некотором классе метрик. В теории Ни [110] t и плоская фоновая метрика η являются нединамическими полями, а скалярное поле ϕ и векторное поле **K** — динамическими. Из этих полей алгебраически строится физическая метрика **g**. Теория содержит три произвольные функции от ϕ и один произвольный параметр; в постньютоновском пределе в случае асимптотически плоской физической метрики (окончательная сшивка с космологическими моделями до сих пор не проведена) имеется четыре произвольных параметра — a, b, d и e. Их можно выбрать так, чтобы ППН-параметры теории совпадали с ППН-параметрами общей теории относительности.

Указанные теории и значения их ППН-параметров кратко представлены в табл. 7. Показано, что все эти теории основываются на инвариантном принципе действия (в основе теории лежит лагранжиан), следовательно, все они являются полуконсервативными, т. е. в них значения параметров α_3, ζ_1, ζ_2, ζ_3 и ζ_4 обращаются в нуль (в табл. 7 они обозначены α_3, ζ_i).

Мы не обсуждали все те разнообразные теории, о которых известно, что они нежизнеспособны. Среди них — класс теорий расслоения, которые предсказывают $\alpha_1 \approx -8$, что сильно противоречит геофизическим данным (п. 3.6), класс «квазилинейных» теорий, в том числе и теория Уайтхеда [111], которые предсказывают ани-

зотропию в локальных измерениях, вызванную распределением масс в Галактике, что сильно противоречит данным по земным приливам (п. 3.6), и класс «квазилинейных» теорий, предсказывающих $\gamma=-1$ в грубом противоречии с измерениями отклонения света и задержки времени (п. 3.3). Мы не обсуждали также те теории, которым еще только предстоит появиться на свет.

3.3. ЭКСПЕРИМЕНТЫ ПО ОТКЛОНЕНИЮ СВЕТА И ЗАДЕРЖКЕ ВРЕМЕНИ: ИССЛЕДОВАНИЕ ЭФФЕКТОВ, СВЯЗАННЫХ С КРИВИЗНОЙ В СОЛНЕЧНОЙ СИСТЕМЕ

Отклонение световых лучей Солнцем и дополнительная задержка времени возвращения сигнала радиолокатора, проходящего вблизи Солнца, определяют ППН-параметр γ. Луч света (или фотон), проходящий мимо Солнца на расстоянии d, отклоняется на угол

$$\delta\theta = \frac{1}{2}(1+\gamma)(4m_\odot/d)[(1+\cos\theta)/2] \qquad (6)$$

([112, 113], ТМЭГ [2], § 7.2), где $m_\odot$ — масса Солнца и θ — угол между линией, соединяющей Землю и Солнце, и направлением приходящего фотона. Для луча, проходящего по касательной к Солнцу, $d=R_\odot$, $\theta\approx 0$ и

$$\delta\theta \approx \frac{1}{2}(1+\gamma)\, 1{,}75''$$

независимо от частоты света. Радиолокационные сигналы, посланные через Солнечную систему к планетам или спутникам, находящимся за Солнцем, возвращаются на Землю, испытав дополнительную неньютоновскую задержку времени, которая дается соотношением ([114], ТМЭГ [2], § 7.2)

$$\delta t \approx \frac{1}{2}(1+\gamma)[250-20\ln(d^2/r)] \text{ мкс},$$

где d — минимальное расстояние, на котором луч проходит от Солнца, выраженное в солнечных радиусах, r — расстояние от планеты или спутника до Солнца, выраженное в астрономических единицах.

Измерения этих двух эффектов дают нам самое точное прямое определение значения параметра γ[1]) на сегодняшний день.

Предсказание отклонения луча света Солнцем явилось одним из величайших успехов эйнштейновской общей теории относительности. Эддингтон, наблюдая солнечное затмение сразу же по окончании первой мировой войны, подтвердил, что видимый свет от звезд действительно отклоняется Солнцем, после чего Эйнштейн стал знаменитым. Однако в экспериментах Эддингтона и его сотруд-

[1]) В последних экспериментах Ризенберга и Шапиро [252] по измерению эффекта задержки величина ошибки не превышает 0,2% от величины самого эффекта. В пределах ошибки измерений не было обнаружено отклонений от величины эффекта, следующего из ОТО.— *Прим. ред.*

ников была достигнута точность всего лишь 30%, и последующие эксперименты были ненамного точнее: разброс результатов составлял от половины до удвоенной величины, предсказанной Эйнштейном, и точности оставались невысокими (см. обзоры [3, 10, 115]). Самое последнее оптическое измерение, проводившееся во время солнечного затмения 30 июня 1973 г., наглядно показывает сложность таких экспериментов. Было получено [116, 117]

$$\frac{1}{2}(1+\gamma)\approx 0{,}95\pm 0{,}11.$$

Невысокая точность объяснялась плохой видимостью из-за пылевой бури непосредственно перед затмением, а также облачностью и дождями во время последующей экспедиции в ноябре 1973 г. В результате число изображений звезд, пригодных для измерений, резко сократилось. Кроме того, масштабы на фотопластинках, полученных во время затмения, к сожалению, отличались от масштабов на пластинках сравнения с изображением опорных звезд. Последние достижения в разработке фотоэлектронных и астрометрических средств, по-видимому, позволяют проводить измерения отклонения света независимо от солнечных затмений [118].

Развитие радиоинтерферометрии с длинными базами в корне изменило эту ситуацию. Методы интерферометрии с длинными и сверхдлинными базами предоставили принципиальную возможность измерения угловых размеров и вариаций углов с точностью лучше чем $3\cdot 10^{-4}$ угловых секунд. Этому техническому прогрессу благоприятствовал ряд «совпадений» на небе: каждый год группы мощных квазизвездных радиоисточников проходят на очень малом угловом расстоянии от Солнца (если смотреть с Земли); сюда входит группа 3C273, 3C279 и 3C48, а также группа 0111+02, 0119+11 и 0116+08. Измеряя относительные отклонения сигналов от квазаров одной и той же группы, радиоастрономы за последние десять лет смогли определить коэффициент $(^1/_2)(1+\gamma)$ в соотношении (6), равный, согласно общей теории относительности, единице. Их результаты приведены на рис. 1.

Одним из главных источников ошибок в этих экспериментах является солнечная корона, отклоняющая радиоволны намного сильнее, чем лучи видимого света, отклонение которых наблюдал Эддингтон. Усовершенствованные методы измерения на двух частотах повысили точности, дав возможность учесть отклонение волн короной, поскольку оно зависит от длины волны и поэтому может быть измерено отдельно от гравитационного отклонения, которое от частоты не зависит. Фомалонт и Шрамек [131] дали полный обзор этих экспериментов и обсудили пути повышения их точности.

Эффект «задержки времени» не был предсказан Эйнштейном; только в 1964 г. этот эффект был обнаружен Шапиро [114] как теоретическое следствие общей теории относительности и других теорий гравитации (см. также [132]). В следующем десятилетии были пред-

приняты попытки измерить этот эффект, воспользовавшись радиолокацией целей, проходящих через точку «дальнего соединения» (цель расположена по другую сторону от Солнца, и сигналы радиолокатора проходят близко от Солнца). Использовались три типа

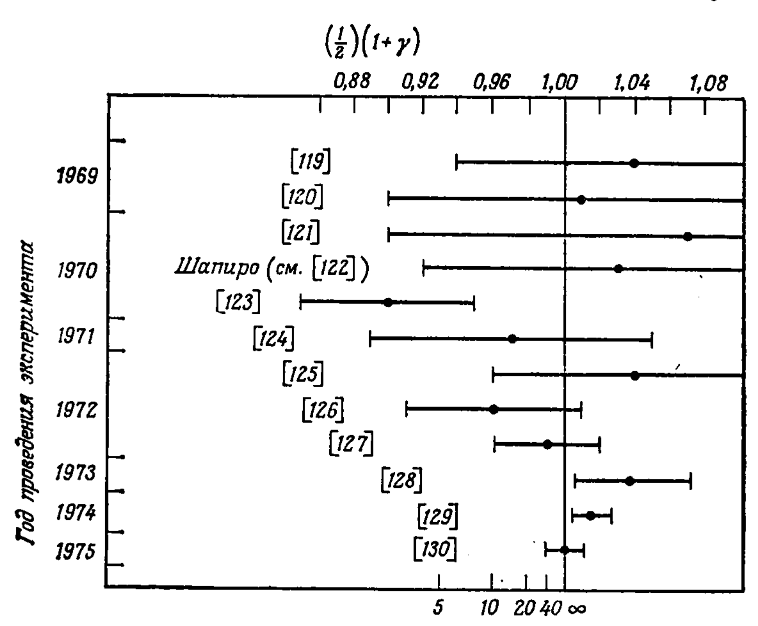


Рис. 1. Результаты экспериментов с 1969 по 1975 г. по отклонению радиоволн.

целей: планеты, такие, как Меркурий и Венера, играли роль пассивных отражателей радарных сигналов; космические корабли, такие, как «Маринер-6» и «Маринер-7», применялись как активные ретрансляторы радиолокационных сигналов, и, наконец, использовались комбинации планеты и космического корабля (космический корабль «на якоре»), такие, как «Маринер-9» на орбите вокруг Марса, орбитальные и спускаемые аппараты «Викинг» и др. Детальный анализ измеряемых задержек времени возвращения сигналов позволил получить значения коэффициента $(^1/_2)(1+\gamma)$, приведенные на рис. 2.

В этом случае, как и при измерении отклонения радиоволн, солнечная корона вносила неопределенности, связанные с замедлением локационных сигналов при прохождении короны; и снова измерения на двух частотах позволили уменьшить ошибки. Другими важнейшими источниками систематических ошибок являются. 1) возмущающие эффекты негравитационных сил, таких, как солнечный ветер, давление излучения и несбалансированные силы,

возникающие в системе контроля за положением космического корабля на траектории; 2) топография планет, не позволяющая определять с помощью пассивной радиолокации положение планет с точностью выше 0,5 км. Использование космических кораблей «на якоре», таких, как орбитальные корабли «Маринер-9» и «Викинг», и применение спускаемых аппаратов должно устранить оба источника указанных ошибок. Действительно, предварительная обработка данных, переданных с «Викинга», дала для величины $(^1/_2)(1+\gamma)$ точность примерно 0,5%, а окончательная точность может достичь 0,1% [138]. Подробные обзоры экспериментальных проблем и планов на будущее см. в работах [139] и [140].

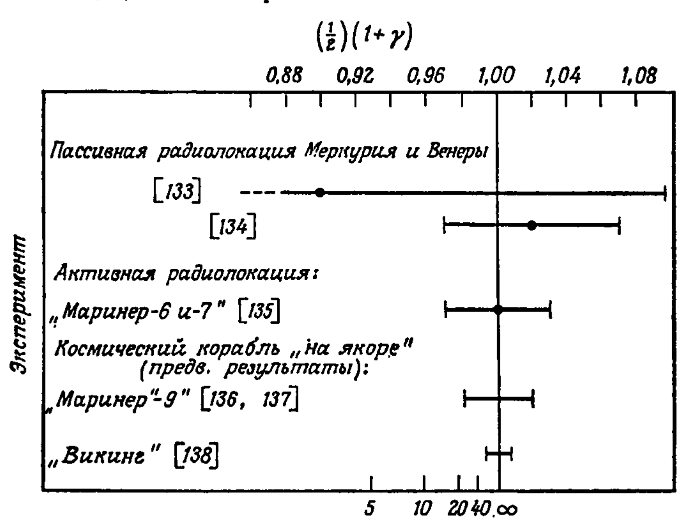


Рис. 2. Результаты экспериментов с 1968 по 1977 г. по задержке времени.

Исходя из этих результатов, мы можем сделать вывод, что коэффициент $(^1/_2)(1+\gamma)$ отличается от единицы не более чем на 1%. В большинстве теорий, представленных в табл. 7, можно таким образом подобрать свободные параметры или граничные космологические условия, что это ограничение будет удовлетворяться. Чтобы коэффициент $(^1/_2)(1+\gamma)$ отличался от единицы не более чем на 1%, в скалярно-тензорных теориях следует принять $\omega > 48$ и $\omega > 23$, если отклонение от единицы не превышает 2%.

3.4. ЛАЗЕРНАЯ ЛОКАЦИЯ ЛУНЫ; ПРОВЕРКА ПРИНЦИПА ЭКВИВАЛЕНТНОСТИ ДЛЯ МАССИВНЫХ ТЕЛ

Слабый принцип эквивалентности (СПЭ) — независимость траекторий пробных тел от их состава — является одним из основных принципов любой метрической теории гравитации (разд. 2.2). Од-

нако этот принцип применим лишь к тем телам, которые по определению обладают пренебрежимо малой внутренней гравитационной энергией. Чтобы описать движение тел, обладающих *конечной* собственной гравитационной энергией, необходимо воспользоваться либо конкретной метрической теорией гравитации, либо ППН-формализмом. В самых первых расчетах, воспользовавшись своей ранней версией ППН-формализма, Нордтведт [88, 141—143] показал, что многие метрические теории гравитации предсказывают нарушение принципа эквивалентности для массивных тел, т. е. утверждают, что ускорение свободного падения зависит от собственной гравитационной энергии. В случае сферически-симметричного тела ускорение из состояния покоя во внешнем гравитационном поле с потенциалом U имеет вид (согласно той версии ППН-формализма, которая представлена в табл. 5) [ср. с уравнением (4)]

$$\mathbf{a} = (m_P/m)\nabla U,$$
$$m_P/m = 1 - \eta\,(E_g/m), \tag{7}$$
$$\eta = 4\beta - \gamma - 3 - \frac{10}{3}\xi - \alpha_1 + \frac{2}{3}\alpha_2 - \frac{2}{3}\zeta_1 - \frac{1}{3}\zeta_2,$$

где E_g — собственная гравитационная энергия тела. Это нарушение принципа эквивалентности для массивных тел известно как «эффект Нордтведта» (возможность такого эффекта впервые отметил Дикки [9], см. также [89, 144]). Этот эффект отсутствует в общей теории относительности ($\eta=0$), но присутствует в теории Бранса — Дикки ($\eta=1/(2+\omega)$). Существование эффекта Нордтведта не влияет на результаты лабораторных экспериментов Этвеша, поскольку для объектов лабораторных размеров $E_g/m \sim 10^{-27}$, т. е. эффект выходит далеко за пределы чувствительности современных экспериментов. Однако для небесных тел отношение E_g/m может быть значительным (10^{-5} для Солнца, 10^{-8} для Юпитера, $4{,}6\cdot10^{-10}$ для Земли), и фактически уже проведен такой «эксперимент Этвеша» с использованием объектов Солнечной системы, который послужил нетривиальной проверкой эффекта Нордтведта. Если бы эффект Нордтведта имел место ($\eta \neq 0$), то Земля должна была падать к Солнцу с ускорением, которое немного отличается от соответствующего ускорения Луны ($E_g/m \sim 0{,}2\cdot10^{-10}$). Это возмущение лунно-земной орбиты приводило бы к поляризации орбиты в направлении к Солнцу по мере того, как Солнце, если смотреть с Земли, движется вокруг системы Земля — Луна. Эта поляризация описывается следующим возмущением расстояния между Землей и Луной:

$$\delta r \approx 9{,}2\eta \cos(\omega_0 - \omega_s)\,t \text{ м}, \tag{8}$$

где ω_0 и ω_s — угловые частоты обращения Луны и Солнца вокруг Земли ([92, 145, 146], подробный вывод этого выражения см. в ТМЭГ [2], § 7.7).

Начиная с августа 1969 г., когда был успешно принят первый лазерный сигнал, отраженный от оптического рефлектора, достав-

ленного на Луну «Аполлоном-11», в соответствии с программой лазерной локации Луны проводились регулярные измерения времени возвращения лазерных импульсов, посылаемых из обсерватории Мак-Дональда в Техасе и отраженных рефлекторами на Луне; измерения проводились с точностью 1 нс (30 см) (см. обзор в работе [147]). Две группы независимо обработали данные, полученные с 1969 по 1975 г.; после строгого учета большого ньютоновского возмущения вида $\delta r \sim 110 \cos(\omega_0 - \omega_s)t$ км, на которое должен был накладываться эффект Нордтведта (8), в обоих случаях эффект Нордтведта не был обнаружен. Результаты двух групп для η таковы:

$$\eta = \begin{cases} 0{,}00 \pm 0{,}03, & \text{(Уильямс и др. [148]),} \\ 0{,}001 \pm 0{,}015, & \text{(Шапиро и др. [149]).} \end{cases}$$

Отсюда ограничение на возможное нарушение СПЭ для массивных тел составляет $7 \cdot 10^{-12}$ (см. табл. 1). Для теории Бранса — Дикки из этих результатов следует ограничение снизу на константу связи ω, равное 29 (результат Шапиро с точностью 2σ).

Повышение точности измерений и усовершенствование теоретического анализа движения Луны может усилить это ограничение на порядок величины [148], в то же время сравнимая точность может быть достигнута при проверке эффекта Нордтведта, если воспользоваться системой Марс — Солнце — Юпитер [149]. Нордтведт [141, 143, 150, 151] рассмотрел другие потенциально наблюдаемые эффекты, связанные с нарушениями принципа эквивалентности.

3.5. СМЕЩЕНИЯ ПЕРИГЕЛИЯ

Объяснение «аномального» смещения перигелия Меркурия, составляющего 43 с за столетие, явилось еще одним из самых ранних успехов общей теории относительности, но и сегодня интерпретация этого эффекта сталкивается с неопределенностью и разногласием. *Измеренное* смещение перигелия точно известно: после учета возмущающих эффектов со стороны других планет оставшееся смещение перигелия известно а) по оптическим наблюдениям Меркурия за последние три столетия с точностью $\sim 1\%$ [152], б) по радиолокационным измерениям за последние десять лет с точностью около 0,5% [149].

Но предсказываемый эффект состоит из двух основных частей: одна обусловлена релятивистской гравитацией, другая — возможной сплюснутостью Солнца, приводящей к искажению гравитационного потенциала Солнца. Для орбиты с большой полуосью a и эксцентриситетом e ППН-формализм предсказывает за один оборот по орбите смещение $\Delta\omega$, равное

$$\Delta\omega = \frac{6\pi m_\odot}{a(1-e^2)} \left\{ \frac{1}{3}(2+2\gamma-\beta) + J_2 \left[\frac{R_\odot^2}{2m_\odot a(1-e^2)} \right] \right\}, \qquad (9)$$

где $R_\odot$ — средний радиус Солнца, J_2 — безразмерная величина солнечного квадрупольного момента, определяемая соотношением $J_2=(C-A)/m_\odot R_\odot^2$, где C и A — моменты инерции относительно оси вращения Солнца и оси, лежащей в экваториальной плоскости, соответственно. Для Меркурия предсказываемое смещение в угловых секундах за столетие равно

$$\dot{\omega}_{☿} = 42{,}95\lambda_p,$$

$$\lambda_p = \left[\frac{1}{3}(2+2\gamma-\beta)+0{,}0003\,(J_2/10^{-7})\right],$$

где величина J_2 представлена нормированной на приближенное значение, которое имело бы место в том случае, если бы Солнце вращалось как однородное твердое тело с наблюдаемой угловой скоростью поверхности [153]. В настоящее время измерения одной только орбиты Меркурия не дают возможности разделить эффекты релятивистской гравитации и солнечного квадрупольного момента, вносящие вклад в величину λ_p. По этой причине в недавней обработке измерений расстояний до Меркурия с помощью радиолокации *предполагалось*, что J_2 имеет величину, соответствующую твердотельному вращению (вклад в λ_p пренебрежимо мал); это позволяло оценить комбинацию ППН-параметров. Результаты имели вид

$$\frac{1}{3}(2+2\gamma-\beta) \approx \begin{cases} 1{,}005 \pm 0{,}020 & \text{(данные 1966—1971 гг., Шапиро и др. [154]),} \\ 1{,}003 \pm 0{,}005 & \text{(данные 1966—1976 гг., Шапиро и др. [138]).} \end{cases}$$

Неопределенность, затруднившая интерпретацию измерений смещения перигелия, возникла из-за серии экспериментов, проведенных в 1966 г. Дикки и Голденбергом (подробный обзор см. в работе [155]). В этих экспериментах измерялась видимая сплюснутость или уплощенность солнечного диска и было обнаружено, что различие видимых полярного и экваториального радиусов составляет $\Delta R = (43{,}3''\pm3{,}3'')\cdot10^{-3}$. При учете сплюснутости поверхностных слоев Солнца, вызванной центробежными силами, этот результат, указывающий на сплюснутость Солнца, можно связать с величиной J_2 (см. [156]):

$$J_2 \approx \frac{2}{3}(\Delta R/R_\odot) - 5{,}3\cdot10^{-6},$$

что дает ($R_\odot$=959″)

$$J_2 \approx (2{,}47\pm0{,}23)\cdot10^{-5} \quad [155].$$

С другой стороны, измерения, проведенные в 1973 г. Хиллом и сотр., дали $\Delta R=(9{,}2''\pm6{,}2'')\cdot10^{-3}$ или

$$J_2 \approx (0{,}10\pm0{,}43)\cdot10^{-5} \quad [157]$$

(см. также [158]). Несоответствие между этими двумя результатами остается невыясненным.

Одна из основных трудностей, возникающих при попытке связать результаты по сплюснутости Солнца с величиной J_2, заключается в необходимости привлекать для этого сложную физику Солнца. Однако существует способ однозначного определения J_2, а именно путем исследования гравитационного поля Солнца на различных расстояниях от него, разделяя тем самым эффект, связанный с J_2, и эффект релятивистской гравитации за счет различной зависимости этих эффектов от расстояния. Одним из методов явилось бы сравнение смещений перигелиев для различных планет. Но смещения перигелиев Венеры, Земли и Марса известны с еще недостаточной точностью, хотя Шапиро и др. [154] подчеркнули, что еще несколько лет радиолокационных наблюдений за внутренними планетами могут сделать такое сравнение возможным. Другой метод — воспользоваться эксцентриситетом орбиты Меркурия ($e \sim 0{,}2$) и искать различные *периодические* возмущения орбиты, вызванные J_2 и релятивистской гравитацией. Чтобы достичь точности, необходимой для таких измерений, потребовалось бы следить за траекторией космического корабля на орбите вокруг Меркурия, но предварительные исследования показали, что J_2 можно было бы определить таким способом с точностью до 10^{-7} [153, 159]. И наконец, космический корабль на сильно вытянутой орбите вокруг Солнца с перигелием в четыре солнечных радиуса позволил бы измерить J_2 с точностью 10^{-8} [53, 153]. Такие программы привели бы, кроме того, к более точному определению γ и β.

3.6. ПРОВЕРКА ЭФФЕКТОВ, СВЯЗАННЫХ С СУЩЕСТВОВАНИЕМ ВЫДЕЛЕННОЙ СИСТЕМЫ ОТСЧЕТА И ВЫДЕЛЕННОГО ПОЛОЖЕНИЯ В ПРОСТРАНСТВЕ

Некоторые теории гравитации предсказывают, что результаты локальных гравитационных экспериментов могут зависеть от скорости лаборатории относительно средней системы покоя Вселенной (эффекты, связанные с существованием выделенной системы отсчета) или от места расположения лаборатории по отношению к соседнему гравитирующему телу (эффекты, связанные с выделенным положением в пространстве). Эффекты, связанные с существованием выделенной системы отсчета, определяются значениями ППН-параметров α_1, α_2, α_3, а эффекты, связанные с выделенным положением в пространстве,— величиной ξ (см. табл. 6). К наиболее важным из этих эффектов относятся: а) вариации и анизотропии в локально измеряемой величине гравитационной постоянной G_L и б) аномальные смещения перигелиев планет.

а. Вариации и анизотропии в величине G_L

Локальное значение гравитационной постоянной может быть определено путем измерения силы притяжения между двумя гравитирующими телами (эксперимент Кавендиша). Однако точность лабораторных экспериментов Кавендиша не очень высока и не позволяет обнаружить постньютоновские эффекты (см. [160]). С другой стороны, высокоточный гравиметр на поверхности Земли — не что иное, как такая разновидность эксперимента Кавендиша, когда достигается особая чувствительность к *вариациям* величины G_L и когда благодаря успехам гравиметрической техники можно искать постньютоновские эффекты. Согласно предсказанию ППН-формализма, G_L имеет вид ([77, 82, 83], ТМЭГ [2], § 6.5)

$$G_L = 1 - \left[4\beta - \gamma - 3 - \zeta_2 - \xi\left(3 + \frac{I}{mR^2}\right)\right] U_{\text{внеш}} + \xi\left(1 - \frac{3I}{mR^2}\right) U_{\text{внеш}} (\mathbf{e}_e \cdot \mathbf{e}_r)^2 + \frac{1}{2}\left[\alpha_3 - \alpha_1 + \alpha_2\left(1 - \frac{I}{mR^2}\right)\right] \mathbf{w}_\oplus^2 - \frac{1}{2}\alpha_2\left(1 - \frac{3I}{mR^2}\right)(\mathbf{w}_\oplus \cdot \mathbf{e}_r)^2,$$

где $U_{\text{внеш}}$ — внешний ньютоновский гравитационный потенциал тела типа Солнца или Галактики, а $\mathbf{e}_e$ — единичный вектор, направленный от Земли к этому телу, $\mathbf{w}_\oplus$— скорость Земли относительно средней системы покоя Вселенной, $\mathbf{e}_r$ — единичный вектор, соединяющий гравиметр с центром Земли, I, m и R — сферический момент инерции Земли, ее масса и радиус соответственно.

За счет эксцентриситета орбитального движения Земли внешний потенциал $U_{\text{внеш}}$, создаваемый Солнцем и Галактикой, варьируется в течение года, но эти вариации составляют всего лишь 10^{-10} и слишком малы, чтобы их можно было надежно обнаружить гравиметрами на Земле; с другой стороны, анизотропный член изменяется с амплитудой, равной $U_{\text{внеш}} \sim U_{\text{галакт}} \sim 5 \cdot 10^{-7}$, благодаря вариации скалярного произведения $(\mathbf{e}_e \cdot \mathbf{e}_r)$ по мере вращения Земли по отношению к вектору $\mathbf{e}_e$. В случае выделенной системы отсчета скорость Земли $\mathbf{w}_\oplus$ складывается из постоянной скорости $\mathbf{w}$ ($\sim$300 км$\cdot$с$^{-1}$) Солнечной системы относительно выделенной системы отсчета и орбитальной скорости движения Земли $\mathbf{v}$ ($\sim$30 км$\cdot$с$^{-1}$) вокруг Солнца. Таким образом, доминирующие вариации величины G_L равны

$$\Delta G_L/G_L \approx \left(\frac{1}{2}\alpha_2 + \alpha_3 - \alpha_1\right)\mathbf{w}\cdot\mathbf{v} - \frac{1}{2}\xi U_{\text{внеш}} (\mathbf{e}_e \cdot \mathbf{e}_r)^2 + \frac{1}{4}\alpha_2[(\mathbf{w}\cdot\mathbf{e}_r)^2 + 2(\mathbf{w}\cdot\mathbf{e}_r)(\mathbf{v}\cdot\mathbf{e}_r)], \quad (10)$$

где уже учтено то обстоятельство, что для Земли $I \approx (^1/_3)mR^2$.

Члены в (10), пропорциональные $\mathbf{w}\cdot\mathbf{v}$ и $(\mathbf{w}\cdot\mathbf{e}_r)(\mathbf{v}\cdot\mathbf{e}_r)$, приводят к годовой вариации величины G_L, что вызывает соответствующее изменение момента инерции Земли, которое в свою очередь приво-

дит к вариациям скорости сидерического вращения Земли Ω. Примем для дальнейшего рассмотрения, что скорость $\mathbf{w}$ равна по величине 300 км·с$^{-1}$ и совпадает по направлению с движением Солнца сквозь космическое микроволновое излучение: $\alpha=165°\pm9°$, $\delta=6°\pm10°$ [161]; тогда амплитуда вариаций равна

$$(\Delta\Omega/\Omega)_{\text{ППН}} \approx 9\left(\frac{2}{3}\alpha_2+\alpha_3-\alpha_1\right)\cdot 10^{-9}.$$

Нордтведт и Уилл [83], а также Рочестер и Смайли [162] показали, что наблюдения вариаций продолжительности дня в совокупности с оценками вклада атмосферного ветра в эти вариации налагает ограничение на предсказываемый эффект, которое дается неравенством

$$\left|\frac{2}{3}\alpha_2+\alpha_3-\alpha_1\right| < 0{,}02.$$

Анизотропные члены в (10) приводят к вариациям гравиметрических измерений и деформаций Земли (если считать ее упругой), которые имеют те же самые характеристики, что и приливы на Земле, если считать ее твердым телом. Предсказываются основные типы приливов со следующими угловыми частотами: Ω (суточные сидерические), 2Ω (полусуточные сидерические) и $2\Omega-\omega$ (полусуточные), где ω — угловая частота движения Земли по орбите. Амплитуды приливов составляют примерно 10^{-8} для приливов, связанных с существованием выделенной системы отсчета, и 10^{-7} для приливов, связанных с выделенным положением в пространстве. Варбуртон и Гудкайнд [163] обработали 18-месячную запись гравиметрических данных, полученных на сверхпроводящем гравиметре (точность измерений соответствовала $\sigma(\Delta G/G)\approx10^{-11}$) на плоскогорье Пиньон в Калифорнии, и установили, что при указанном выше выборе $\mathbf{w}$ эти данные позволяют наложить как на α_2, так и на ξ верхние ограничения, даваемые неравенствами

$$|\alpha_2| < 10^{-3}, \quad |\xi| < 10^{-3},$$

что на порядок лучше грубых ограничений, полученных Уиллом [82]. Другие эффекты, связанные с этими вариациями величины G_L, рассмотрены Нордтведтом и Уиллом [83], см. также ТМЭГ [2], § 7.4.

6. Аномальные смещения перигелиев

Эффекты, связанные с существованием выделенной системы отсчета и с выделенным положением в пространстве, могут приводить к следующим аномальным смещениям перигелиев для планет, выраженным в угловых секундах за столетие ([77, 83] и ТМЭГ [2], § 7.3):

$$\begin{aligned}\dot{\omega}_{☿} &\approx -104\alpha_1+70\alpha_2+12\cdot10^4\alpha_3+63\xi,\\ \dot{\omega}_{\oplus} &\approx -165\alpha_1+9\alpha_2+21\cdot10^5\alpha_3+14\xi.\end{aligned} \qquad (11)$$

Сравнивая эти аномальные смещения с измеренными смещениями для Меркурия и Земли и учитывая ограничения на параметры, налагаемые геофизическими данными, из (11) получаем ограничение на α_3 и тем самым ограничение на α_1, которые даются неравенствами

$$|\alpha_3| < 2\cdot 10^{-6}, \quad |\alpha_1| < 2\cdot 10^{-2}.$$

Данные по лазерной локации Луны также могут быть использованы для наложения ограничений на возмущения орбиты Луны, вызванные существованием выделенной системы отсчета [83, 145].

3.7. ЭКСПЕРИМЕНТ С ГИРОСКОПОМ И ПРОВЕРКА ЭФФЕКТА УВЛЕЧЕНИЯ ИНЕРЦИАЛЬНЫХ СИСТЕМ ОТСЧЕТА

Начиная с 1960 г. много усилий было направлено на проверку теорий гравитации с помощью сверхпроводящего гироскопа на орбите [87, 164—166]. Цель эксперимента состоит в том, чтобы измерить прецессию оси вращения гироскопа относительно удаленных звезд по мере движения гироскопа по орбите вокруг Земли. Два релятивистских эффекта являются потенциально наблюдаемыми. Одним из этих эффектов является «геодезическая» прецессия, в которую дает вклад кривизна пространства вокруг Земли. В случае полярной орбиты, если ось гироскопа лежит в плоскости орбиты, геодезическая прецессия приводит к вековому повороту оси вращения с амплитудой

$$\text{Геодезическая прецессия} \sim \frac{1}{3}(1+2\gamma)\cdot(7''\cdot\text{год}^{-1}).$$

Другой эффект — прецессия Ленза — Тирринга, или «увлечение инерциальных систем отсчета», вызванное вращением Земли (детальное обсуждение см. в МТУ [1], § 19.2 и 33.4). В случае полярной орбиты, если ось вращения гироскопа перпендикулярна плоскости орбиты, увлечение инерциальных систем отсчета приводит к вековому повороту оси вращения с амплитудой

$$\text{Прецессия Ленза — Тирринга} \sim \frac{1}{8}(4\gamma+4+\alpha_1)\cdot(0{,}05''\cdot\text{год}^{-1}). \quad (12)$$

Цель эксперимента — достичь точности 10^{-3} угловых секунд за год. Подробный вывод формул, касающихся эксперимента с гироскопом, приводится в работе [167], МТУ [1], § 40.7, и в ТМЭГ [2], § 6.4.

Вариант эксперимента с гироскопом был недавно предложен Ван Паттеном и Эвериттом [168], когда «гироскопом» являлся сам спутник на орбите вокруг Земли. Увлечение инерциальных систем отсчета приводит к повороту плоскости орбиты вокруг оси, параллельной оси вращения Земли. В случае спутника на полярной орбите на высоте 800 км над Землей скорость поворота составляет $\sim(^1/_8)(4\gamma+4+\alpha_1)\cdot(0{,}18''\cdot\text{год}^{-1})$. Чтобы исключить эффекты поворота, обусловленные другими причинами (например, вследствие

квадрупольного момента Земли), необходимо иметь два спутника, которые вращаются в противоположных направлениях по почти одинаковым орбитам. Согласно оценке Ван Паттена и Эверитта, с помощью спутников, свободных от сноса, можно достичь точности эксперимента 1%, воспользовавшись данными, передаваемыми с орбиты в течение двух с половиной лет.

3.8. ЛАБОРАТОРНЫЕ ЭКСПЕРИМЕНТЫ ПО ПРОВЕРКЕ ПОСТНЬЮТОНОВСКОЙ ГРАВИТАЦИИ

Поскольку гравитационные силы очень слабы, большинство проверок постньютоновских эффектов в Солнечной системе предполагает использовать в качестве источников гравитации Солнце и планеты. Один из недостатков таких экспериментов состоит в том, что экспериментатор не имеет возможности осуществлять контроль над источниками и, следовательно, не может манипулировать с экспериментальной конфигурацией, чтобы проверять и повышать чувствительность аппаратуры или чтобы повторять эксперимент. Несмотря на этот недостаток экспериментов с объектами в масштабах Солнечной системы, малость постньютоновских поправок препятствует разумным лабораторным экспериментам, но есть одно исключение.

Этим исключением является эксперимент Кройцера [169], в котором проводится сравнение активной и пассивной гравитационных масс фтора и брома. В эксперименте Кройцера использовались весы Кавендиша для сравнения ньютоновской гравитационной силы, создаваемой цилиндром из тефлона (76 вес. % фтора), и силы, создаваемой таким количеством жидкой смеси трихлорэтилена и дибромметана (74 вес. % брома), которое обладает такой же пассивной гравитационной массой, что и указанный цилиндр (а именно массой жидкости, вытесненной цилиндром, который в ней не тонет и не всплывает). Главный вывод Кройцера состоял в том, что эти силы оказались одинаковыми и, следовательно, отношения активной массы (m_A) к пассивной массе (m_P) для фтора и брома одинаковы с точностью $5\cdot 10^{-5}$ [169], т. е.

$$\left|\frac{(m_A/m_P)_{\text{фтор}}-(m_A/m_P)_{\text{бром}}}{(m_A/m_P)_{\text{бром}}}\right| < 5\cdot 10^{-5}. \tag{13}$$

(Более подробно об эксперименте Кройцера см. в работах [170, 171].)

Уилл [91] проанализировал этот эксперимент, воспользовавшись тем вариантом ППН-формализма, в котором рассматриваются заряженные точечные массы, и показал, что электростатическая энергия связи тела E_e дает вклад в m_A/m_P, который имеет вид

$$m_A/m_P = 1 + \frac{1}{2}\zeta_3\,(E_e/m_P). \tag{14}$$

Поскольку ядра фтора и брома содержат различные относительные количества электростатической энергии связи, соотношения (13)

и (14) дают следующее ограничение на параметр ζ_3:

$$|\zeta_3| < 6 \cdot 10^{-2}.$$

Это является обобщением и уточнением результата, полученного ранее Торном и др. [13].

Недавно Брагинский и др. [172] высказали предположение, что технический прогресс, возможно, позволит в ближайшем десятилетии провести ряд лабораторных постньютоновских экспериментов. Этот прогресс, благодаря которому такие эксперименты станут осуществимы, связан с разработкой чувствительных систем с очень низким уровнем диссипации (таких, как системы крутильных маятников с нитями из плавленого кварца или сапфира, охлажденных до температур $\leqslant 0{,}1$ K), с созданием массивных диэлектрических кристаллов, охлажденных до миллиградусов Кельвина, и микроволновых резонаторов со сверхпроводящими стенками. К экспериментальным возможностям, перечисленным в работе [172], относятся измерение спин-спинового взаимодействия двух вращающихся тел, поиск временны́х вариаций гравитационной постоянной (разд. 6), поиск эффектов, связанных с существованием выделенной системы отсчета и выделенного положения в пространстве, и измерение увлечения инерциальных систем отсчета вращающимся телом.

3.9. ВЫВОДЫ

Общая теория относительности победоносно прошла через все проверки постньютоновских гравитационных эффектов в Солнечной системе. Но то же самое можно сказать о еще нескольких альтернативных теориях, поэтому ясно, что Солнечная система не может быть единственной ареной сопоставления гравитационной теории и эксперимента. Поэтому, чтобы усилить это сопоставление, необходимо обратиться к предсказаниям теории, выходящим за рамки постньютоновского круга явлений. Именно к этому новому кругу явлений — к гравитационному излучению, к проверкам в звездных системах и к космологическим тестам — мы сейчас и переходим.

4. ГРАВИТАЦИОННОЕ ИЗЛУЧЕНИЕ КАК ИНСТРУМЕНТ ДЛЯ ПРОВЕРКИ ГРАВИТАЦИОННОЙ ТЕОРИИ

Проверки постньютоновских гравитационных эффектов в Солнечной системе сыграли важную роль при выявлении тех метрических теорий гравитации, которые оказались нежизнеспособными, и тех теорий, которые могут быть жизнеспособными. Однако такие эксперименты не дают полной картины, поскольку в них исследуется лишь малая часть, постньютоновский предел, всей совокупности предсказаний гравитационных теорий. Это усугубляется тем обстоятельством, что имеется несколько метрических теорий гравитации,

которые формулируются совершенно по-разному и все же могут полностью совпадать с общей теорией относительности в постньютоновском пределе и тем самым находиться в согласии с современными проверками в Солнечной системе (см. табл. 7). Таким образом, центр тяжести проблемы проверки этих теорий переносится с постньютоновского приближения на новые области «пространства предсказаний», в основу проверки кладутся новые возможные явления.

Новую почву для проверок дает гравитационное излучение [1]). Недавние исследования показали, что метрические теории гравитации могут отличаться друг от друга и от общей теории относительности в своих предсказаниях свойств гравитационных волн по трем главным вопросам: 1) они могут предсказывать различные состояния поляризации для типичных гравитационных волн (п. 4.1); 2) они могут предсказывать отличие скорости слабых гравитационных волн от скорости света (п. 4.2); 3) они могут предсказывать различные мультипольности, т. е. монопольность, дипольность, квадрупольность и т. д. гравитационного излучения, испущенного данными источниками (п. 4.3). Для использования поляризации и скорости в качестве проверок требуется регулярное детектирование гравитационного излучения — перспектива, по-видимому, почти неосуществимая. С другой стороны, мультипольность гравитационных волн можно исследовать, анализируя обратное влияние гравитационного излучения на источник (реакцию излучения) для различных мультиполей. Примером может служить изменение периода орбит в системе двух тел, вызванное изменением энергии системы в результате испускания гравитационного излучения. Такая проверка, по-видимому, уже возможна в случае двойного пульсара PSR 1913+16 (см. пп. 5.2, а и 5.2, б).

4.1. ПОЛЯРИЗАЦИЯ ГРАВИТАЦИОННЫХ ВОЛН

Общая теория относительности предсказывает для слабого гравитационного излучения два независимых состояния поляризации: моды «+» и «×» на языке МТУ (§ 35.6) или состояния спиральности +2 и —2 на языке квантовой теории поля. Однако общая теория относительности является, вероятно, единственной теорией, делающей такое предсказание: любая другая из известных жизнеспособных метрических теорий гравитации предсказывает для типичной гравитационной волны более чем две поляризации. Действительно, в самом общем случае слабой гравитационной волны теория может предсказать *шесть* составляющих мод поляризации, выражаемых через шесть «электрических» компонент тензора Римана R_{0i0j}, которые определяют вынуждающие силы, действующие на детектор [173, 174].

[1]) С обзором предсказаний интенсивности различных астрофизических источников гравитационного излучения читатель может познакомиться в статье Торна [253].— *Прим. ред.*

В случае слабой плоской гравитационной волны, распространяющейся по оси z, амплитуды шести мод поляризации можно записать в виде двух действительных функций $\Psi_2(u)$, $\Phi_{22}(u)$ и двух комплексных функций $\Psi_3(u)$ и $\Psi_4(u)$ от запаздывающего времени $u=t-z$ (обозначения, принятые в формализме Ньюмена — Пенроуза). Эти функции даются соотношениями

$$\Psi_2 = -\frac{1}{6} R_{z0z0}, \quad \Psi_3 = \frac{1}{2}(-R_{x0z0} + iR_{y0z0}),$$

$$\Psi_4 = R_{y0y0} - R_{x0x0} + 2iR_{x0y0}, \quad \Phi_{22} = -(R_{x0x0} + R_{y0y0}).$$

На рис. 3 показано действие каждой из мод на сферу из пробных тел; Ψ_4 и Φ_{22} являются чисто поперечными, Ψ_2 — чисто продольным,

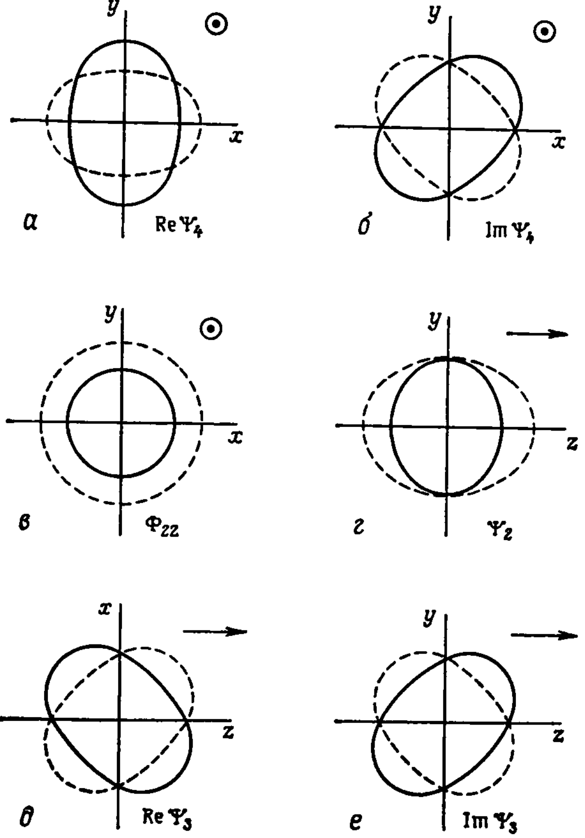


Рис. 3. Шесть мод поляризации слабой, плоской, нулевой гравитационной волны, допустимые в любой теории гравитации.

Показаны смещения, вызываемые каждой модой на сфере, состоящей из пробных частиц. Волна распространяется в направлении оси $+z$ и зависит от времени как $\cos \omega t$. Сплошной линией показана картина в момент $\omega t = 0$, а пунктирной линией — в момент $\omega t = \pi$. В плоскости, перпендикулярной рисунку, смещения отсутствуют.

а Ψ_3 — смесь поперечного и продольного полей. Общая теория относительности допускает только две Ψ_4-моды, тогда как скалярно-тензорные теории допускают моды Ψ_4 и Φ_{22}. Теории можно классифицировать в соответствии с числом и типом мод, которые эти теории допускают для типичной волны. Например, в наиболее общем классе, II_6, $\Psi_2 \neq 0$, могут присутствовать все шесть мод; в классе III_5 $\Psi_2 \equiv 0$, $\Psi_3 \neq 0$ и могут присутствовать пять мод. В этих двух классах амплитуды ненулевых мод зависят от наблюдателя, т. е. они не инвариантны относительно преобразований Лоренца, оставляющих неизменным волновой вектор [«малая группа» $E(2)$]. В классе N_3 $\Psi_2 \equiv \Psi_3 \equiv 0$, и присутствуют три моды (Ψ_4, Φ_{22}); в N_2 $\Psi_2 \equiv \Psi_3 \equiv \equiv \Phi_{22} \equiv 0$, и присутствует две моды (Ψ_4); в O_1 $\Psi_2 \equiv \Psi_3 \equiv \Psi_4 \equiv 0$, и присутствует всего одна мода (Φ_{22}). В этих трех классах амплитуда имеющихся мод инвариантна относительно «малой группы»; только в классах (N_2, N_3, 0_1) возможно описывать волны с помощью состояний с определенной спиральностью (± 2 для Ψ_4, 0 для Φ_{22}). В табл. 8 показано, к каким классам относятся жизнеспособные на сегодняшний день теории гравитации, обсуждавшиеся в п. 3.2.

Поскольку в самом общем случае волна определяется шестью электрическими компонентами тензора Римана, то соответствующим образом спроектированный гравитационно-волновой детектор может однозначно определять шесть амплитуд (см. рис. 3) и тем самым класс волны, *если только известно направление, в котором находится источник*. Направление должно определяться либо с помощью гравитационно-волновой интерферометрии, либо благодаря связи гравитационной волны с оптическим явлением, таким, например, как Сверхновая, либо с помощью какого-нибудь другого метода. Если направление неизвестно, то можно сузить возможный класс волны, но нельзя определить ее класс однозначно (наблюдаемые: 6 вынуждающих сил; неизвестны: 6 амплитуд и 2 направляющих косинуса). Подробная стратегия определения или сужения класса волн разобрана в работе [174]. Если предположить *á priori*, что класс волны — это N_2, N_3 или 0_1 (волны с определенной спиральностью), то с помощью одного подходящего детектора можно одновременно определить и класс волны, и ее направление [175, 176]. Если наблюдаемый класс волны является более общим (присутствует большее число поляризаций) по сравнению с тем, что предсказывает данная теория, то такая теория является нежизнеспособной; однако если наблюдаемый класс менее общий, чем предсказываемый теорией, из этого нельзя сделать определенного вывода, поскольку специфика астрофизического источника (например, его симметрия) может приводить к отсутствию некоторого состояния поляризации, которое имело бы место в типичной ситуации.

Таблица 8

Свойства гравитационного излучения в жизнеспособных на сегодняшний день теориях гравитации

Теория	Класс $E(2)$	Имеется ли определенная спиральность?	ПМ-параметры $\varkappa_1$	ПМ-параметры $\varkappa_2$	Дипольный параметр $\varkappa_D$	Знак энергии	$c_g = c_{эм}$?	Может ли теория совпадать с ОТО в ППН-приближении?
Общая теория относительности	N_2	Да (± 2)	12	11	0	+	Да	
Скалярно-тензорные теории:								
БВН — Бекенштейна	N_3	Да ($\pm 2,0$)	$12 - \frac{5}{2+\omega}$	$11 - \frac{45}{8+4\omega}$	$\frac{2}{2+\omega} \times [1 + 4\Lambda(2+\omega)]^2$	+	Да	Нет
Бранса — Дикки	N_3	Да ($\pm 2,0$)	$12 - \frac{5}{2+\omega}$	$11 - \frac{45}{8+4\omega}$	$\frac{2}{2+\omega}$	+	Да	Нет
Векторно-тензорные теории:								
Уилла — Нордтведта	III_5	Нет	[a]	[a]	[a]	[a]	Нет	Да
Биметрические теории:								
Розена	II_6	Нет	$-21/2$	$-23/2$	$-20/3$	+/−	Нет	Да
Лейтмана — Ли	II_6	Нет	$-21/2$	$73/8$	$-125/3$	+/−	Нет	Да
Расталла	II_6	Нет	[a]	[a]	[a]	[a]	Нет	Да
Теории расслоения:								
Ни	II_6	Нет	-18	-19	$-400/3$	+/−	Нет	Да

[a] Вычисления по определению этих величин еще не проведены.

4.2. СКОРОСТЬ ГРАВИТАЦИОННЫХ ВОЛН

Некоторые теории гравитации предсказывают скорость распространения гравитационных волн c_g, равную скорости света $c_{эм}$ (см. табл. 8), тогда как, согласно другим теориям, скорости должны различаться, и в случае слабых гравитационных полей, как правило,

$$(c_g - c_{эм})/c_{эм} \sim U/c_{эм}^2,$$

где U — локальный ньютоновский гравитационный потенциал. Внутри Галактики недалеко от нас или в поле скопления галактик в созвездии Девы $U/c_{эм}^2 \sim 10^{-7}$, поэтому, сравнивая времена прихода света и гравитационных волн от отдельного события типа Сверхновой в скоплении галактик в созвездии Девы, можно было бы наложить ограничение на эту разницу скоростей [173]:

$$|c_g - c_{эм}|/c_{эм} < 10^{-9} \cdot (\text{точность измерения временно́го запаздывания})/(1\ \text{неделя}).$$

4.3. МУЛЬТИПОЛЬНОСТЬ ГРАВИТАЦИОННЫХ ВОЛН; ИЗЛУЧЕНИЕ ОТ ДВОЙНЫХ СИСТЕМ

Как известно, согласно общей теории относительности, наинизшим мультиполем в гравитационном излучении является квадруполь в том смысле, что если разложить гравитационное поле на мультиполи в волновой зоне, воспользовавшись для этого тензорными сферическими гармониками, то в разложении будут фигурировать только $\overleftrightarrow{T}^{E2}_{lm}$ и $\overleftrightarrow{T}^{B2}_{lm}$ с $l \geqslant 2$ (обозначения и подробное обсуждение см. в работе [177]), т. е. поперечные, бесследовые гармоники со спином 2. Обращаясь к материальным источникам гравитационного излучения и подходящим образом определяя мультипольные моменты распределения вещества и гравитационного поля внутри ближней неволновой зоны, окружающей источник, предыдущее утверждение можно сформулировать иначе [177]: самым низшим мультиполем источника, способным генерировать гравитационное излучение, является квадруполь. В случае источников, в которых происходят медленные движения и гравитационное поле слабо, таких, как системы двойных звезд, квадрупольное излучение действительно преобладает над всеми другими мультиполями. В двойной системе с полной массой $m = m_1 + m_2$, с приведенной массой $\mu = m_1 m_2/m$, с текущим расстоянием между компонентами r и скоростью орбитального движения v квадрупольное излучение приводит к следующему темпу потери орбитальной энергии [178]:

$$\frac{dE}{dt} = -\left\langle \frac{\mu^2 m^2}{r^4} \frac{8}{15} (12v^2 - 11\dot{r}^2) \right\rangle, \tag{15}$$

где $\dot{r} = dr/dt$, а угловые скобки означают усреднение по орбите. По третьему закону Кеплера, эта потеря энергии приводит к умень-

шению орбитального периода P:

$$\dot{P}/P = -\frac{3}{2} E^{-1}\, dE/dt.$$

(Однако на этот счет имеется и противоположная точка зрения [179].) Квадрупольное излучение приводит также к уменьшению углового момента системы и к соответствующему уменьшению эксцентриситета орбиты (библиографию и сводку формул см. в работе [180]). Как подчеркнул Фолкнер [181], эффекты квадрупольного излучения могут играть важную роль в эволюции двойных систем со сверхкоротким периодом. Но наиболее обещающей проверкой существования квадрупольного излучения будут, по всей вероятности, наблюдения за изменениями периода $\dot{P}$ в двойном пульсаре (п. 5.2, а).

Однако в отличие от общей теории относительности почти все альтернативные теории гравитации предсказывают наряду с квадруполем и высшими мультиполями присутствие в гравитационном излучении всех мультиполей, включая монополь и *диполь* [182—184]. В случае двойных звездных систем учет вклада от дополнительных мультиполей приведет к двум изменениям в формуле для темпа потери энергии (15): а) изменятся численные коэффициенты в (15) и б) появится дополнительный член (из-за дипольных моментов), который зависит от собственной гравитационной энергии связи звезд. Окончательная формула для dE/dt может быть записана в форме, куда входят безразмерные параметры, значения которых зависят от конкретной изучаемой теории. Два параметра, $\varkappa_1$ и $\varkappa_2$, называют ПМ-параметрами, поскольку они относятся к той части dE/dt, которая соответствует результату Петерса и Мэтьюса [178] для общей теории относительности. Параметр $\varkappa_D$ относится к дипольному вкладу за счет собственной гравитации. Окончательно имеем [184]

$$\frac{dE}{dt} = -\left\langle \frac{\mu^2 m^2}{r^4} \left[\frac{8}{15} (\varkappa_1 v^2 - \varkappa_2 \dot{r}^2) + \frac{1}{3} \varkappa_D \mathfrak{S}^2 \right] \right\rangle, \tag{16}$$

где $\mathfrak{S}$ — разница в собственной гравитационной энергии связи между двумя объектами в двойной системе, отнесенная к единице массы (более строгое определение $\mathfrak{S}$ см. в работах [182, 183]).

В табл. 8 приведены предсказываемые значения $\varkappa_1$, $\varkappa_2$ и $\varkappa_D$ для тех теорий, в которых эти значения рассчитаны. Мы отмечаем странный результат, получаемый во всех теориях, за исключением скалярно-тензорных: дипольное излучение уносит отрицательную энергию, т. е. увеличивает энергию системы ($\varkappa_D < 0$), а ПМ-излучение может уносить либо положительную, либо отрицательную энергию в зависимости от теории и характера орбиты. Можно было бы утверждать (и заранее ясно, что некоторые будут так утверждать), что уже только это является достаточным основанием для суждения о нежизнеспособности таких теорий. Однако это лишь теоретическое ограничение на теории, не имеющее под собой в случае гравитации

почти никакой экспериментальной основы, поэтому мы сосредоточим внимание на наблюдательных данных, свидетельствующих за или против такого странного эффекта.

Единственная из теорий, представленных в табл. 8, которая автоматически предсказывает полное отсутствие дипольного гравитационного излучения,— это общая теория относительности. В скалярно-тензорных теориях типа теории ВВН — Бекенштейна также можно избежать дипольного излучения, например, если функция $\omega(\phi)$ дается соотношением $\omega(\phi)={}^1/_2(4-3\phi)/(\phi-1)$, тогда $1+4\Lambda(2+\omega)=0=\varkappa_D$. В этом случае оказывается, что локально измеренная гравитационная постоянная G_L является истинной постоянной и что теория предсказывает отсутствие эффекта Нордтведта($4\beta-\gamma-3=0$) [185]. Для других теорий, упомянутых в табл. 8, возможно, имеется тесная связь между существованием или отсутствием дипольного излучения, знаком энергии, уносимой гравитационными волнами, и классом теории по отношению к группе $E(2)$. Общая теория относительности предсказывает волны наиболее узкого класса N_2 (если не считать тривиальный класс 0_1) и с определенной спиральностью (± 2). Это следует из того обстоятельства, что точные эйнштейновские уравнения поля могут быть записаны в виде (МТУ [1], § 20, [186])

$$\Box\theta^{\mu\nu}=-16\pi\tau^{\mu\nu}, \quad \tau^{\mu\nu}{}_{,\nu}=0,$$

где величина $\theta^{\mu\nu}$ связана с пространственно-временно́й метрикой, $\tau^{\mu\nu}$ — псевдотензор энергии-импульса вещества и гравитационных полей и координаты здесь выбраны так (выбор калибровки), чтобы удовлетворялось равенство $\theta^{\mu\nu}{}_{,\nu}=0$. Однако имеется дополнительная калибровочная свобода, допускаемая сохранением $\tau^{\mu\nu}$ ($\tau^{\mu\nu}{}_{,\nu}=0$), с помощью которой, преобразуя $\theta^{\mu\nu}$, нельзя исключить лишь два спиральных состояния. На теоретико-полевом языке волны обладают определенной спиральностью, поскольку они связаны с сохраняющимся током. Но, кроме того, именно закон сохранения $\tau^{\mu\nu}{}_{,\nu}=0$ исключает вклад в $\theta^{\mu\nu}$ от монопольного и дипольного моментов любых сколь угодно релятивистских источников. Поэтому нет, по-видимому, ничего удивительного, если теории гравитации, в которых имеются волны более общего класса, чем N_2, предсказывают дипольное излучение. В чисто динамических теориях по крайней мере одно из дипольных динамических полей не связано с сохраняющимся током (в противном случае возникали бы состояния с определенной спиральностью), и поэтому дипольное излучение оказывается допустимым. Исключения составляют скалярно-тензорные теории (N_3), в которых скалярное поле автоматически обладает определенной нулевой спиральностью, несмотря на то, что оно может быть связано с несохраняющимся током в источнике [184]. В теориях с первичной геометрией существование нединамических полей нарушает калибровочную инвариантность динамических полей, дающих вклад в физическую метрику, и тем самым допускает ненулевой

вклад монопольного и дипольного моментов источника. Поэтому нет, по-видимому, ничего удивительного и в том, что такие теории предсказывают отрицательный или неопределенный знак излучаемой энергии, подобно тому как на языке квантовой теории поля понятия определенной спиральности, квантуемости и положительной определенности энергии неотделимы друг от друга.

Для систем, содержащих компактные объекты типа нейтронных звезд ($\mathfrak{S} \sim E_g/m \sim 0{,}5$), дипольный вклад в (16) может преобладать над ПМ-вкладом ($v^2 \sim \dot{r}^2 \sim 10^{-6}$ при $v \sim 200$ км·с$^{-1}$), если $\varkappa_D \neq 0$. Наиболее перспективным полем деятельности по проверке существования дипольного гравитационного излучения может служить двойной пульсар (п. 5.2, б). Такая проверка может служить решающим свидетельством за или против широкого класса гравитационных теорий, которые неотличимы от общей теории относительности при проверках, проводимых в Солнечной системе.

5. ПРОВЕРКА ПО ЗВЕЗДНЫМ СИСТЕМАМ: ДВОЙНОЙ ПУЛЬСАР

Летом 1974 г. Рассел Халс и Джозеф Тейлор проводили систематический обзор неба в поисках новых пульсаров на обсерватории Аресибо в Пуэрто-Рико. В ходе этого обзора они зарегистрировали 50 пульсаров, причем 40 из этих пульсаров до этого были неизвестны, и провели целый ряд наблюдений, в том числе измерили периоды их импульсов с точностью до 1 мкс. Однако один из этих пульсаров под номером PSR 1913+16 оказался особенным: кроме того, что он обладал периодом 59 мс, т. е. самым коротким периодом среди всех пульсаров после пульсара в Крабовидной туманности, все попытки измерить этот период с точностью ± 1 мкс оказались безуспешными, поскольку пульсар обнаруживал явные изменения периода: в разные дни различие достигало 80 мкс, а иногда период менялся на 8 мкс за 5 мин [187]. Такое поведение весьма несвойственно пульсарам, и Халс и Тейлор быстро сделали вывод, что наблюдаемые изменения периода являются результатом доплеровских смещений вследствие орбитального движения пульсара вокруг компаньона. В конце сентября 1974 г. Халс и Тейлор получили точную кривую скорости этой «спектральной двойной с единственной линией». Согласовав во всех деталях эту кривую с кеплеровской орбитой двух тел, они получили следующие элементы орбиты системы: K_1, полуамплитуду вариации радиальной скорости пульсара относительно центра масс системы; P_b, орбитальный период двойной системы, скорректированный с учетом движения обсерватории; e, эксцентриситет орбиты; ω, долготу периастра в данную эпоху (сентябрь 1974 г.); $a_1 \sin i$, проекцию большой полуоси орбиты пульсара, где i — угол наклона орбиты по отношению к плоскости, касательной к небесной сфере, и, наконец, $f_1 = (m_2 \sin i)^3/(m_1 + m_2)^2$, функцию масс, где m_1 и m_2 — массы пульсара и его компаньона. Кроме того, они

получили период пульсара P_p в системе покоя пульсара, введя необходимые поправки на орбитальные доплеровские смещения в данную эпоху (сентябрь 1974 г.). Эти результаты приведены в средней колонке табл. 9 [187].

Таблица 9

Параметры PSR 1913+16

Параметр	Значение, полученное из данных по периоду (лето 1974 г.)	Значение, полученное из данных по времени прихода импульсов (сентябрь 1974 г.—октябрь 1975 г.)
K_1, км·с$^{-1}$	199 ± 5	
P_b, с	$27\,908\pm7$	$27906{,}980\pm0{,}002$
e	$0{,}615\pm0{,}010$	$0{,}61717\pm0{,}00005$
ω_2, град	179 ± 1	$178{,}861\pm0{,}007$
$a_1\sin i$, см	$(6{,}96\pm0{,}13)\cdot10^{10}$	$(7{,}0043\pm0{,}0004)\cdot10^{10}$
f_1, $m_\odot$	$0{,}13\pm0{,}01$	$0{,}13126\pm0{,}00002$
P_p, с	$0{,}059030\pm1$	$0{,}059029995272\pm5$
Литература	Халс и Тейлор [187]	Тейлор и др. [188]

Однако в конце сентября 1974 г. наблюдатели перешли к другой методике наблюдений, которая позволяла существенно повысить точность. В этой методике измеряются времена прихода индивидуальных импульсов (а не период, т. е. разница времени прихода соседних импульсов) и затем из этих времен вычитаются времена прихода, предсказываемые с помощью наиболее стабильного пульсара и с использованием параметров орбиты [189]. Затем параметры орбиты уточняются за счет обработки полученных разностей времен прихода методом наименьших квадратов. Результаты такой обработки данных, полученных к октябрю 1975 г., приведены в правой колонке табл. 9 [188].

Открытие PSR 1913+16 вызвало сильное оживление среди релятивистов [а о том, что происходило в редакции «Письма в Астрофизический журнал» (Astrophysical Journal Letters), нечего и говорить], поскольку стало ясно, что вновь открытая система может стать новой лабораторией для исследования релятивистской гравитации. Постньютоновские орбитальные эффекты достигли бы здесь величин порядка $v^2\sim K_1^2\sim5\cdot10^{-7}$, $m/r\sim f_1/a_1\sin i\sim3\cdot10^{-7}$, т. е. имели бы на порядок более высокие значения, чем соответствующие значения для Меркурия, а малость орбитального периода ($\sim$8 ч) приводила бы к усилению всех вековых эффектов, таких, как, например, смещение периастра. Эти надежды подтвердились, когда в декабре 1974 г. Тейлор сообщил [190] о том, что измеренное смещение периастра оказалось равным $4{,}0\pm1{,}5$ град·год$^{-1}$ (сравните с Меркурием!). Более того, эта система, по всей видимости, должна

быть «чистой» лабораторией, не подверженной таким сложным астрофизическим процессам, как перетекание масс. Радиосигнал пульсара никогда не испытывал затмение компаньоном, что налагало ограничение на геометрический размер компаньона, и дисперсия импульсного радиосигнала не проявляла каких-либо заметных изменений в ходе движения пульсара по орбите, что указывало на отсутствие в системе плотной плазмы, которая бы там имелась в случае перетекания массы от компаньона к пульсару. Эти данные практически исключают в качестве компаньона звезду главной последовательности, хотя и можно разумным образом подобрать такую звезду, удовлетворив при этом геометрическим ограничениям по отсутствию затмений и по измерениям дисперсии, но такая звезда приводила бы к слишком большому смещению периастра ($>$5000 град$\cdot$год$^{-1}$), связанному с приливными деформациями компаньона в гравитационном поле пульсара [191, 192]. В качестве еще одного кандидата на роль компаньона рассматривали гелиевую звезду главной последовательности, которая могла бы удовлетворить как геометрическим ограничениям, так и ограничениям по смещению периастра. Однако оценки расстояния до пульсара (5 кпс [187]) и ослабления излучения вдоль луча зрения ($\sim$3,3 звездные величины [193]) указывали на то, что такая гелиевая звезда была бы на две звездные величины ярче, чем наблюдательные ограничения ($m_v \gtrsim 23$, $m_R \gtrsim 21$) на любой оптический объект, связанный с пульсаром [194, 195]. Другими возможными кандидатами в компаньоны являются плотные звездные объекты: белый карлик, нейтронная звезда или черная дыра.

Попытки получить более точное представление о свойствах компаньона потребовали обратиться к конструктивным сценариям образования и эволюции этой системы (табл. 10). Наиболее предпочтительным сценарием, по-видимому, является эволюция, начавшаяся с фазы рентгеновской двойной системы; при этом конечным продуктом этой фазы явились две нейтронные звезды [192, 196, 200]. Однако были построены и альтернативные сценарии, приводящие в качестве компаньона к белому карлику [197, 200], черной дыре [192, 200, 201] или гелиевой звезде [192, 200]. Природа компаньона весьма существенна для различных релятивистских и астрофизических эффектов в системе, к обсуждению которых мы переходим.

Двойной пульсар вызвал оживление среди релятивистов, поскольку он может играть две совершенно различные роли: 1) он является «чистой» лабораторией с потенциально сильными релятивистскими эффектами, в которой можно проверять гравитационную теорию (п. 5.2), и 2) *это первая из всех известных систем, для которой релятивистская теория гравитации может быть использована в качестве практического метода определения астрофизических параметров* (п. 5.1). Мы начнем именно с этого второго пункта, учитывая новую и совершенно неожиданную роль релятивистской гравитации.

Таблица 10

Библиография по двойному пульсару

А. Наблюдения и астрофизика	
Наблюдения пульсара	Другие наблюдения
Халс и Тейлор, 1975 [187]	Хьельминг и Гибсон, 1975 [205]
Тейлор, 1975 [190]	Ханан и др., 1975 [206]
Тейлор и др., 1976 [188]	Дэвидсен и др., 1975 [193]
Сценарии эволюции	Бернакка и др., 1975 [207]
Фленнери и ван ден Хойвел, 1975 [196]	Кристиан и др., 1976 [194]
Веббинк, 1975 [192]	Насер и др., 1977 [208]
Ван Хорн и др., 1975 [197]	Ван Цитерс и Рибски, 1977 [209]
Ван ден Берг, 1975 [198]	Методы обработки данных
Уилер, 1976 [199]	Уилер, 1975 [210]
Смарр и Блэндфорд, 1976 [200]	Блэндфорд и Теукольски, 1975 [211], 1976 [189]
Бисноватый-Коган и Комберг, 1976 [201]	Эпштейн, 1977 [212]
Астрофизические эффекты	Природа компаньона
Озерной и Шишов, 1975 [202]	Мастерс и Робертс, 1975 [191]
Бальбус и Бречер, 1976 [203]	Озерной и Рейнгард, 1975 [213]
Шапиро и Терзиан, 1976 [204]	Робертс и др., 1976 [195]
Б. Релятивистская гравитация	
Общая теория относительности	Смещения периастра
Дамур и Руффини, 1974 [214]	Уилл, 1975 [221]
Бречер, 1975 [215]	Баркер и О'Коннелл, 1975 [222]
Брумберг и др., 1975 [216]	Нордтведт, 1975 [219]
Эспозито и Харрисон, 1975 [217]	Прецессия оси вращения пульсара
Демьянский и Шакура, 1976 [218]	Баркер и О'Коннелл, 1975 [222]
Масса пульсара, определяемая по красному смещению	Зельдович и Шакура, 1975 [223]
Уилер, 1975 [210]	Хари Дасс и Радхакришнан, 1975 [224]
Блэндфорд и Теукольски, 1975 [211], 1976 [189]	Рудольф, 1977 [225]
Проверки гравитационной теории	Гравитационное излучение
Эрдли, 1975 [182]	Эрдли, 1975 [182]
Нордтведт, 1975 [219]	Вагонер, 1975 [180], 1976 [226]
Уилл, 1976 [220], 1977 [184]	Уилл и Эрдли, 1977 [183]
Уилл и Эрдли, 1977 [183]	Уилл, 1977 [184]

5.1. РЕЛЯТИВИСТСКАЯ ГРАВИТАЦИЯ КАК МЕТОД ИЗМЕРЕНИЯ АСТРОФИЗИЧЕСКИХ ПАРАМЕТРОВ

а. Сдвиг периастра

Согласно ППН-формализму, темп прецессии периастра двух тел с массами m_1 и m_2 дается соотношением [ср. (9)]

$$\frac{d\omega}{dt}=\frac{6\pi m}{P_b a(1-e^2)}\left[\frac{1}{3}(2+2\gamma-\beta)+\frac{1}{6}(2\alpha_1-\alpha_2+\alpha_3+2\zeta_2)\frac{m_1 m_2}{m^2}\right], \quad (17)$$

где $m=m_1+m_2$ — суммарная масса, P_b — орбитальный период. Второй член в (17) может быть значительным только в том случае,

когда ни одно из тел не имеет массу, пренебрежимо малую по сравнению с массой другого тела (для Меркурия и Солнца $m_1 m_2/m^2 \sim \sim 2 \cdot 10^{-7}$), однако так или иначе этот член отсутствует в любой полностью консервативной теории гравитации ($\alpha_1 = \alpha_2 = \alpha_3 = \zeta_2 = 0$). В случае почти кеплеровской орбиты, описываемой соотношением $P_b/2\pi = a^{3/2} m^{-1/2}$, формулу (17) можно записать в следующем виде:

$$\frac{d\omega}{dt} = \frac{3m^{2/3}}{(P_b/2\pi)^{5/3}(1-e^2)}\left[\frac{1}{3}(2+2\gamma-\beta) + \frac{1}{6}(2\alpha_1 - \alpha_2 + \alpha_3 + 2\zeta_2)\frac{X}{(1+X)^2}\right],$$

где $X = m_1/m_2$. P_b и e для двойного пульсара известны (табл. 9), поэтому для $d\omega/dt$ в градусах за год имеем

$$\frac{d\omega}{dt} = 2{,}11^\circ\left(\frac{m}{m_\odot}\right)^{2/3}\left[\frac{1}{3}(2+2\gamma-\beta) + \right.$$
$$\left. + \frac{1}{6}(2\alpha_1 - \alpha_2 + \alpha_3 + 2\zeta_2)\frac{X}{(1+X)^2}\right] \text{год}^{-1}.$$

Поскольку смещение периастра зависит от неизвестной полной массы m и от отношения масс X, воспользоваться этим эффектом для проверки гравитационной теории невозможно. Вместо этого оказывается более выгодным обратить таблицу: задаться какой-нибудь конкретной теорией гравитации и воспользоваться измеренным смещением периастра для определения полной массы системы. Например, общая теория относительности предсказывает смещение ($\gamma = \beta = 1$, $\alpha_1 = \alpha_2 = \alpha_3 = \zeta_2 = 0$), равное

$$(d\omega/dt)_{\text{ОТО}} = 2{,}11^\circ (m/m_\odot)^{2/3} \text{ год}^{-1}.$$

Тейлор и др. [188] измерили смещение периастра, которое оказалось равным

$$(d\omega/dt)_{\text{измер}} \approx 4{,}22 \pm 0{,}04^\circ \text{ год}^{-1},$$

отсюда полная масса системы $m \approx 2{,}83 \pm 0{,}04\ m_\odot$. Это первый известный нам случай, когда общая теория относительности была использована как практический метод проведения астрономического измерения с высокой точностью ($\sim 1\%$).

Однако прежде чем измеренную по смещению периастра массу объявить окончательным результатом, следует сделать два предостережения.

1. Могут существовать нерелятивистские источники смещения периастра в системе — это главным образом возможные квадрупольные деформации компаньона [аналогичные эффекту квадрупольного момента Солнца, отраженному в формуле (9)]. Эти деформации бывают двух типов: приливные деформации, вызванные гравитационным полем пульсара (эти деформации существенны только в том случае, если компаньоном является гелиевая звезда), и вращательные деформации, которые могут приводить к смещению перигелия либо в направлении вращения орбиты, либо в противоположном направлении в зависимости от ориентации оси вращения компань-

она (эти деформации существенны, если компаньоном является гелиевая звезда или быстро вращающийся белый карлик). Если же компаньоном является черная дыра, нейтронная звезда или невращающийся белый карлик, то имеется только релятивистская прецессия периастра (подробное обсуждение этого вопроса см. в работах [195, 200, 221]).

2. В альтернативных теориях гравитации, подобных представленным в табл. 7, ППН-параметры могут находиться в согласии с ППН-параметрами общей теории относительности, и тем не менее предсказываемое ими смещение может заметно отличаться от смещения, даваемого выражением (17). Хотя орбитальное движение можно описывать в постньютоновском приближении ($m/r \sim v^2 \ll 1$), структуру пульсара ($m/R \sim 0{,}5$), вообще говоря, необходимо описывать точными релятивистскими уравнениями гравитационной теории (примеры моделей нейтронных звезд в альтернативных теориях см. в работах [103, 182, 183, 227—229]). В результате могут возникать большие релятивистские поправки к массам — активной, пассивной, инертной и т. д., которые проявляются в ньютоновских и постньютоновских уравнениях движения. В общей теории относительности независимо от того, насколько релятивистским является источник, никаких таких поправок не возникает. В биметрической теории Розена, например, третий закон Кеплера и предсказание относительно смещения периастра принимают следующий вид [183]:

$$(P_b/2\pi) = \mathfrak{G}^{-1/2} a^{3/2} m^{-1/2},$$
$$d\omega/dt = 6\pi \mathfrak{P} \mathfrak{G}^{-1} m / P_b a (1 - e^2),$$

где $\mathfrak{G} = 1 - 4s_1 s_2/3$, $\mathfrak{P} = (1 - \frac{8}{9}(s_1 + s_2) + 28 s_1 s_2/27)\mathfrak{G} + 0(s^3)$, а s_1 и s_2 связаны с собственными гравитационными энергиями связи, отнесенными к единице массы двух тел [$s_1 - s_2 = \mathfrak{S}$ в формуле (16)]. Для нейтронных звезд в теории Розена величина s может достигать 0,6. В результате масса, определяемая по смещению периастра, дается соотношением

$$m \approx (2{,}83 \pm 0{,}04)\, m_\odot \mathfrak{G}^2 \mathfrak{P}^{-3/2}.$$

Например, если компаньон является нейтронной звездой с массой, близкой к массе пульсара, то $m \approx 6{,}96\, m_\odot$ [183].

Приняв к сведению эти предостерегающие замечания, мы продолжим наше обсуждение, считая, что полная масса системы близка к 2,83 солнечных масс.

б. Гравитационное красное смещение и квадратичное доплеровское смещение: метод взвешивания пульсара

Релятивистские гравитационные эффекты могут служить методом, позволяющим провести первое точное измерение массы нейтронной звезды. На наблюдаемый период пульсара влияет не только проекция его орбитальной скорости на луч зрения (линейное доп-

леровское смещение), но и абсолютная величина этой скорости (квадратичное доплеровское смещение). Эти эффекты описываются формулой

$$(P_p)_{\text{набл}}/(P_p)_{\text{изл}} = 1 + \mathbf{n}\cdot\mathbf{v}_1 + \frac{1}{2}v_1^2 + m_2/r,$$

где $\mathbf{n}$ — единичный вектор, направленный вдоль луча зрения, $\mathbf{v}_1$ — скорость пульсара и r — расстояние между пульсаром и компаньоном. Выразив это отношение через элементы кеплеровской орбиты, его можно записать следующим образом:

$$(P_p)_{\text{набл}}/(P_p)_{\text{изл}} = 1 + K_1[\cos(\omega+\phi) + e\cos\omega] + B\cos\phi + C, \quad (18)$$

где ϕ — азимутальный угол в плоскости орбиты, отсчитываемый от положения периастра, C — неизмеряемая постоянная. Коэффициенты K_1 (см. табл. 9) и B даются соотношениями

$$K_1 = \frac{2\pi a_1 \sin i}{P_b(1-e^2)^{1/2}}, \qquad B = \frac{m_2^2(m_1+2m_2)e}{m^2 a_1(1-e^2)}$$

(заметим, что $B/K_1 \sim [m/a_1]^{1/2} \sim v$). Чтобы однозначно определить величины m_1, m_2, a_1 и $\sin i$, измерение B можно было бы скомбинировать с измеренными значениями функции масс f_1, проекции большой полуоси $a_1 \sin i$ и полной массы m, измеренной по смещению периастра. Если, например, $m=2{,}83\, m_\odot$, то масса пульсара будет равна

$$m_1 = 1{,}42\,[3-(1+B/5\cdot10^{-7})^{1/2}]\, m_\odot.$$

Но измерение B было бы абсолютно невозможным, если бы не было релятивистского смещения периастра. Именно смещение периастра ($\omega=\omega_0+\dot{\omega}t$) позволяет разделить в уравнении (18) член с амплитудой K_1 и B-член; если бы не это смещение, два члена в (18) были бы абсолютно неразличимы. Это объясняется зависимостью линейного доплеровского смещения от ориентации орбиты по отношению к направлению луча зрения; эта ориентация меняется по мере прецессии периастра, тогда как квадратичные эффекты доплеровского красного смещения зависят только от неизменных параметров орбиты. Поскольку периастр поворачивается на 4,22 градуса в год, полное разделение этих эффектов будет возможно по прошествии примерно 23 лет. Однако, как показали Блэндфорд и Теукольски [189, 211], практически измерения времени прихода импульсов с точностью 1 мс, проводимые систематически на протяжении 5 лет, позволили бы измерить B с точностью 10% (измерение B с точностью 1% потребовало бы около 10 лет). Такие измерения массы нейтронной звезды являются важным источником данных при проверке уравнений состояния вещества при высоких плотностях.

в. Прецессия оси вращения пульсара

Если пульсар является быстро вращающейся нейтронной звездой, то он должен испытывать такие же эффекты прецессии оси сво-

его вращения, как и гироскоп на орбите вокруг Земли (п. 3.7). Доминирующим эффектом является геодезическая прецессия, определяемая соотношением [222, 224, 225]

$$\langle d\mathbf{n}/dt\rangle = \mathbf{\Omega}\times\mathbf{n},$$

$$\mathbf{\Omega} = (3\pi/P_b)\left[m_2^2/ma(1-e^2)\right]\left[\frac{1}{3}(2\gamma+1)+\frac{2}{3}\left(\gamma+1+\frac{1}{4}\alpha_1\right)\times\right.$$

$$\left.\times(m_1/m_2)\right]\mathbf{n}_0,$$

где γ, α_1 — ППН-параметры, $\mathbf{n}$ — единичный вектор, параллельный оси вращения пульсара, $\mathbf{n}_0$ — единичный вектор нормали к плоскости орбиты. Для двойного пульсара величина $\mathbf{\Omega}$ равна примерно одному градусу за год (ср. с гироскопом на орбите вокруг Земли, п. 3.7); однако отметим, что никакой прецессии не происходит, если ось вращения пульсара перпендикулярна плоскости орбиты.

Но если прецессия все-таки происходит, ее можно использовать для проверки ППН-параметров; однако может оказаться более плодотворным воспользоваться релятивистской прецессией для исследования природы механизма излучения пульсара. По мере прецессии пульсара луч зрения наблюдателя пересекает поверхность нейтронной звезды на различных широтах (таким образом можно получать двумерную информацию о диаграмме направленности испускаемого луча, а также исследовать широтные вариации спектра и поляризации [200]). К сожалению, в большинстве моделей пульсара излучаемые радиоимпульсы имеют карандашную диаграмму направленности, поэтому излучение пульсара может временами совсем пропадать.

5.2. ИСПОЛЬЗОВАНИЕ ДВОЙНОГО ИМПУЛЬСА ДЛЯ ПРОВЕРКИ РЕЛЯТИВИСТСКОЙ ГРАВИТАЦИИ

а. Проверка существования гравитационного излучения

Согласно предсказанию общей теории относительности, квадрупольное гравитационное излучение вызывает уменьшение энергии двойной системы согласно (15). В случае кеплеровской орбиты формула (15) принимает вид

$$\frac{dE}{dt} = -\frac{32}{5}\left(\frac{\mu}{m}\right)^2\left(\frac{m}{a}\right)^5\left(1+\frac{73}{24}e^2+\frac{37}{96}e^4\right)(1-e^2)^{-7/2}.$$

Из величин измеренных параметров двойного пульсара с необходимостью следует изменение орбитального периода, равное [180, 226]

$$\frac{\dot{P}_b}{P_b} = -3{,}8\cdot 10^{-9}\,\frac{(m/2{,}83m_\odot)^{4/3}}{\sin i}\left[1-0{,}36\,\frac{(m/2{,}83m_\odot)^{-1/3}}{\sin i}\right]\text{год}^{-1}.$$

Предсказываемое изменение периода мало чувствительно к углу наклона, если только этот угол не лежит в интервале от 20 до 30° ($0<m_1<0{,}8\,m_\odot$). Блэндфорд и Теукольски [189] показали, что

ошибка в определении $\dot{P}_b/P_b$, если проводить измерения в течение T лет, ведет себя примерно как $10^{-7}\,T^{-3/2}$ при условии, что времена прихода импульсов могут измеряться с точностью 10^{-3} с и фазовая когерентность импульсов поддерживается на протяжении всего этого времени. Таким образом, за 10 лет можно измерить $\dot{P}_b/P_b$ с точностью 10%, причем за это время предполагается определить индивидуальные массы компонентов и угол наклона, что в свою очередь сделает предсказание об изменении периода более определенным.

Подтверждение этого эффекта выходит за рамки проверки одного лишь факта существования гравитационного излучения. Этим проверяется и то, уносит ли гравитационное излучение энергию из системы. Недавно Розен [230] сделал следующее утверждение: принимая во внимание симметрию по времени уравнений поля для гравитационного излучения во всех теориях гравитации, можно предположить, что решения этих уравнений также должны быть симметричными по времени, т. е. решение должно быть суммой запаздывающих и опережающих волн одинаковой амплитуды в противоположность общепринятым запаздывающим, направленным от источника волновым решениям. Такие симметричные решения также описывают гравитационное излучение (R_{0i0j}), измеряемое лабораторными детекторами, но эти решения не связаны с переносом энергии. Обнаружение $\dot{P}_b/P_b$ исключило бы такую возможность [1]).

Однако прежде чем рассматривать этот эффект как надежную проверку существования гравитационного излучения, следует учесть другие возможные причины изменений периода. Поскольку все проверки гравитационной теории, обсуждаемые в этой главе, апеллируют к обнаружимым изменениям как орбитального, так и собственного периодов пульсара, мы дадим здесь обзор причин изменений обоих этих периодов (табл. 11).

Приливная диссипация. Приливы, возникающие на компаньоне из-за гравитационного поля пульсара, должны приводить к изменению орбитальной и вращательной энергии из-за нагрева за счет вязкости (соответствующими приливами, возникающими на пульсаре, можно пренебречь вследствие его малого размера). Темп изменения периода, обусловленного этой диссипацией, дается соотношением [200, 220]

$$\dot{P}_b/P_b \approx -3\cdot 10^{-13} X\,(1+X)^2 \left(1-\frac{P_b}{\tau}\right)\left(\frac{R}{10^5\ \text{км}}\right)^9 \left(\frac{m}{2{,}83 m_\odot}\right)^{-11/3} \times$$
$$\times \langle\mu\rangle_{13}\ \text{год}^{-1},$$

где $\langle\mu\rangle_{13}$ — усредненный коэффициент вязкости в единицах 10^{13} г $\times$ $\times$ см$^{-1}\cdot$с$^{-1}$, R и τ — радиус и период собственного вращения компаньона соответственно. Отсутствие затмений при наблюдениях

[1]) Согласно измерениям Вайсберга и Тейлора [254], изменение периода обращения двойного пульсара подтверждает формулу для мощности гравитационного излучения ОТО с погрешностью около 15%.— *Прим. ред.*

Таблица 11

Сравнение вековых эффектов в периодах двойного пульсара

Эффект	$[\dot{P}/P]$пульсара, год$^{-1}$	$[\dot{P}/P]$орбит, год$^{-1}$
Проверки гравитационной теории		
Гравитационное излучение		
Квадрупольное, согласно общей теории относительности	Пренебрежимо мало	$2\cdot10^{-9}$
Дипольное	Пренебрежимо мало	$3\cdot10^{-7}$
Нарушение законов сохранения	$4\cdot10^{-7}$	$4\cdot10^{-7}$
Конкурирующие эффекты		
Приливная диссипация		
Нейтронная звезда или черная дыра в качестве компаньона	Пренебрежимо мало	Пренебрежимо мало
Белый карлик, гелиевая звезда в качестве компаньона	Пренебрежимо мало	Неопределенно [a]
Потеря массы	Неопределенна	$< 10^{-14}$
Ускорение, связанное с распределением масс в Галактике	$2\cdot10^{-13}$	$2\cdot10^{-13}$
Наблюдаемые значения (2σ)	$(4{,}6\pm0{,}2)\cdot10^{-9}$ (Тейлор и др., 1976 [188]	$(3{,}6\pm0{,}7)\cdot10^{-9}$ (Тейлор, сообщение на «Техасском» симпозиуме в Мюнхене, 1978)

[a] См. предшествующее обсуждение.

сигнала от пульсара указывает на то, что $R<10^5$ км [187]. Если причиной диссипации является молекулярная вязкость, то $\langle\mu\rangle_{13}<$ $<10^{-10}$ и приливная диссипация вызывает крайне незначительные изменения орбитального периода. Однако если компаньоном пульсара является белый карлик или гелиевая звезда, то вызванные приливными силами макроскопические сдвиговые движения [231] или турбулентность в атмосфере [203] могли бы приводить к $\langle\mu\rangle_{13}\gtrsim1$. Магнитная вязкость в белом карлике могла бы приводить к $\langle\mu\rangle_{13}\sim$ $\sim10^5$ [200].

Вращение Галактики. Поскольку Галактика вращается неоднородно, относительная скорость Солнечной системы и двойной системы меняется вдоль луча зрения как функция времени. Возникающее в результате переменное доплеровское смещение приводит к наблюдаемым изменениям как собственного, так и орбитального периодов пульсара; однако оценки месторасположения и расстояния до пульсара вместе с общепринятым законом вращения Галактики дают [204, 220]

$$|\dot{P}_b/P_b|=|\dot{P}_p/P_p|\sim2\cdot10^{-13}\ \text{год}^{-1}.$$

Потеря энергии. Испускание из пульсара энергии в различных формах (частицы, электромагнитное излучение и т. д.) приводит к уменьшению кинетической энергии вращения пульсара и тем самым к увеличению его периода:

$$dE/dt = -I(2\pi/P_p)^2 \dot{P}_p/P_p,$$

где dE/dt — темп потери энергии, I — момент инерции пульсара. Орбитальный период тоже меняется, темп этого изменения равен

$$\dot{P}_b/P_b = -\frac{1}{2}\dot{m}_1/m,$$

где m_1 — масса пульсара. Однако поскольку $dE/dt \leqslant dm_1/dt$, мы можем написать

$$\dot{P}_b/P_b \leqslant 1\cdot 10^{-6}(m/2{,}83m_\odot)^{-1} I_{45}(\dot{P}_p/P_p),$$

где $I_{45} = I/10^{45}$ г·см^2. Поскольку наблюдаемая величина $\dot{P}_p/P_p$ меньше чем 10^{-8} год$^{-1}$ (табл. 11), то величина $\dot{P}_b/P_b$ за счет потери энергии должна быть меньше 10^{-14} год$^{-1}$.

б. Проверка существования дипольного гравитационного излучения

Как уже отмечалось в п. 4.3, большинство теорий гравитации, отличных от общей теории относительности, предсказывают существование дипольного гравитационного излучения. Поскольку величина этого эффекта в двойной системе зависит от собственных гравитационных энергий связи двух тел, то двойная система — это идеальный объект для проверки существования дипольного излучения. Согласно общей теории относительности, энергии связи нейтронных звезд могут достигать половины от их масс покоя, а в других теориях эти энергии связи еще больше, так что эффект дипольного излучения, если он имеет место, приводил бы к более быстрым изменениям периода, чем общерелятивистский эффект квадрупольного излучения.

Предсказываемый темп изменения периода, усредненный по орбите [ср. с (16)], определяется соотношением [182, 184]

$$\dot{P}_b/P_b = -\varkappa_D \mu \mathfrak{S}^2 (2\pi/P_b)^2 \left(1 + \frac{1}{2}e^2\right)(1-e^2)^{-5/2}.$$

Для двойного пульсара это дает

$$\dot{P}_b/P_b = -(3{,}1\cdot 10^{-7})\varkappa_D(\mathfrak{S}/0{,}1)^2(\mu/m_\odot) \text{ год}^{-1}.$$

Верхний предел на $\dot{P}_b/P_b$, полученный из наблюдений, равен 10^{-7} год$^{-1}$ (табл. 11). Воспользовавшись этим фактом, можно установить предел для параметра $\varkappa_D$ и тем самым проверить альтернативные теории, перечисленные в табл. 8, но вследствие неопределенности масс компонент и неопределенности в отождествлении

компаньона пульсара решающая проверка в настоящее время невозможна. Однако на рис. 4 показаны грубые ограничения, которые можно было бы наложить на $\varkappa_D$ при различных массах пульсара, предполагая, что компаньоном пульсара является 1) нейтронная звезда или 2) белый карлик. Значения величины $\mathfrak{S}$ рассчитывались с помощью общерелятивистских моделей нейтронной звезды [184]. Расчеты, на основании которых построен рис. 4, были проведены в постньютоновском приближении [184], и поэтому в них невозможно

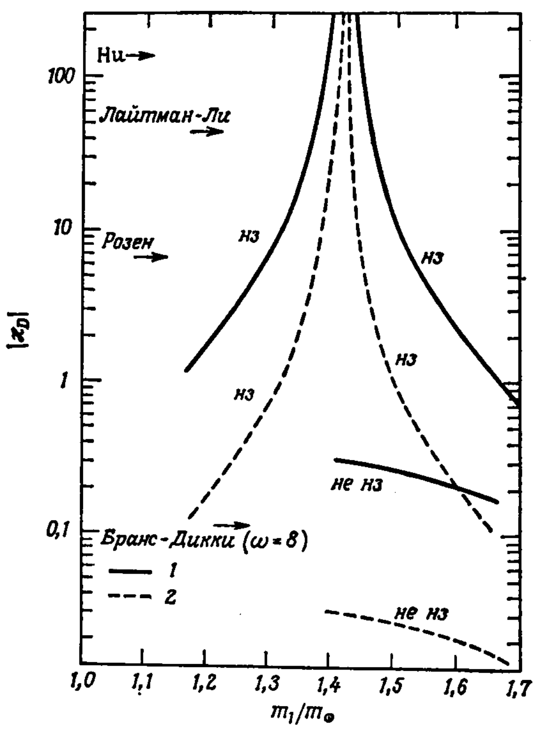


Рис. 4. Ограничения на дипольный параметр $|\varkappa_D|$.

Получены из верхних наблюдательных ограничений (в 2σ) на величину $|\dot{P}_b/P_b|$, равных 10^{-7} год$^{-1}$ (сегодняшнее значение) и 10^{-8} год$^{-1}$ (предполагаемое значение в будущем), ограничения приведены в зависимости от массы пульсара. Рассматриваются два возможных компаньона пульсара; нейтронная звезда (кривая «нз») и не нейтронная звезда (кривая «не нз»). Предполагается, что компаньон вносит незначительный вклад в смещение периастра ($m = 2{,}83\ m_\odot$) и в уменьшение радиуса орбиты. Предсказываемые значения $|\varkappa_D|$ указаны стрелками. 1 — $|\dot{P}_b/P_b| < 10^{-7}$ лет$^{-1}$, 2 — $|\dot{P}_b/P_b| < 10^{-8}$ лет$^{-1}$.

было учесть экстремально релятивистскую структуру пульсара, следовательно, приведенные ограничения на параметр $\varkappa_D$ следует считать лишь некоторым приближением. Расчеты, в которых точно учитывалась релятивистская структура пульсара, были проведены в рамках биметрической теории Розена Уиллом и Эрдли [183].

Основные различия результата этих расчетов по сравнению с предыдущими состоят в следующем: 1) возможность бо́льших масс ($6{,}96 m_\odot$ в случае двух одинаковых нейтронных звезд), определяемых по смещению периастра, 2) бо́льшие значения величины $\mathfrak{S}$ для моделей нейтронных звезд в теории Розена.

В ходе дальнейших наблюдений системы, усилив ограничение на $\dot{P}_b/P_b$, установив природу и массу компаньона, можно будет провести решающие проверки тех теорий, которые без этого можно было считать жизнеспособными альтернативами общей теории относительности.

в. Проверка постньютоновских законов сохранения

Некоторые метрические теории гравитации предсказывают возможность «самоускорений» двойной системы в направлении периастра с ускорением, равным [220]

$$\mathbf{a}_{CM} = \frac{\pi m_1 m_2 (m_1 - m_2)\, e}{P m^{3/2} a^{3/2} (1-e^2)^{3/2}} (\alpha_3 + \zeta_2)\, \mathbf{n},$$

где $\mathbf{n}$ — единичный вектор, направленный от центра масс системы к точке периастра массы m_1. В любой полуконсервативной теории гравитации $\alpha_3 = \zeta_2 = 0$ и такой эффект отсутствует, следовательно, этот эффект может иметь место только в тех теориях гравитации, в которых нарушаются глобальные постньютоновские законы сохранения импульса. Однако современные эксперименты в Солнечной системе (разд. 3) налагают ограничение только на параметр α_3, а именно $|\alpha_3| < 2 \cdot 10^{-5}$. На параметр ζ_2 нет ни одного сильного ограничения, хотя Шапиро и Теукольски [232], воспользовавшись данными по гравитационному красному смещению для белых карликов, получили очень слабое ограничение $|\zeta_2| < 100$. Другие эксперименты типа лазерной локации Луны (п. 3.4) дают проверку для комбинаций ППН-параметров, включающих ζ_2.

В настоящее время известно несколько жизнеспособных теорий гравитации, в которых $\zeta_2 \neq 0$. Любая теория, основанная на инвариантном принципе наименьшего действия, автоматически является полуконсервативной [78]; все теории, перечисленные в табл. 7, попадают в этот класс теорий. И все же ненулевое значение ζ_2 может иметь важные физические следствия; являясь параметром «закона сохранения», ζ_2 в некотором смысле служит мерой того, «насколько сильное гравитационное поле генерируется гравитационной энергией». Вагонер и Мелоун [233] показали, что вклад собственной гравитационной энергии в активную гравитационную массу нейтронной звезды чувствителен к значению параметра ζ_2, и призывали к осторожности при попытках выяснить природу компактного объекта (т. е. отличить нейтронную звезду от черной дыры), основанных только на определении активной гравитационной массы по кеплеровским орбитам. Таким образом, верхнее ограничение на

аномальное ускорение центра масс явилось бы прямой проверкой законов сохранения в гравитации и позволило бы дать ограничение на ППН-параметр ζ_2.

В двойном пульсаре наблюдаемый эффект, связанный с таким аномальным ускорением, проявляется в виде вековых изменений как наблюдаемого периода пульсара, так и наблюдаемого орбитального периода (эти изменения включают в себя и доплеровское смещение), равные

$$\frac{\dot{P}_p}{P_p}=\frac{\dot{P}_b}{P_b}\approx -4\cdot 10^{-7}\,\frac{X(1-X)}{(1+X)^2}\left(\frac{m}{2{,}83m_\odot}\right)^{2/3}\times$$
$$\times(\alpha_3+\zeta_2)\left(\frac{t}{1\ \text{год}}\right)\text{год}^{-1},\quad t\leqslant 10\ \text{лет},$$

где t — время с сентября 1974 г., когда стали учитывать влияние смещения периастра на направление вектора $\mathbf{n}$.

Наблюдения Тейлора и др. [188] вплоть до сентября 1975 г. указывают на то, что для самого пульсара

$$\dot{P}_p/P_p\approx(4{,}6\pm 0{,}2)\cdot 10^{-9}\ \text{год}^{-1},$$

Тем самым ограничение, налагаемое на ζ_2, составляет

$$|\zeta_2|<0{,}1\,[(1+X)^2/8X(1-X)]\,(2{,}83m_\odot/m)^{2/3}.$$

Поскольку предсказываемый эффект растет линейно со временем, повторяемые измерения величины $\dot{P}_p/P_p$ будут усиливать ограничение на ζ_2 (в предположении, что масса пульсара уже измерена и поэтому само предсказание не изменится). Наиболее важным конкурирующим астрофизическим эффектом является потеря энергии пульсаром, причем предполагается, что этот эффект может являться причиной конечного значения $4{,}6\cdot 10^{-9}$ год$^{-1}$ для величины $\dot{P}_p/P_p$.

В принципе эти два эффекта можно разделить, поскольку они по-разному зависят от времени: если темп потери энергии примерно постоянен, то $\dot{P}_p/P_p$ будет константой, тогда как если имеется вековое ускорение центра масс, то $\dot{P}_p/P_p$ будет в современную эпоху изменяться в грубом приближении линейно со временем.

Проблема векового ускорения центра масс двойной системы имеет любопытную историю. Леви-Чивита [234] первым отметил, что общая теория относительности *предсказывает* вековое ускорение в направлении периастра орбиты, и отыскал некоторую двойную систему в качестве кандидата, для которого, как он полагал, этот эффект станет в какой-то момент наблюдаемым. Эддингтон и Кларк повторили расчет [235], воспользовавшись полученными Де Ситтером [236] уравнениями движения для n тел. Сначала они получили вековое ускорение противоположного знака по сравнению с результатом Леви-Чивита, затем они обнаружили ошибку в уравнениях Де Ситтера и пришли в конце концов к выводу, что вековое ускоре-

ние равно нулю. Робертсон [237] независимо пришел к тому же выводу, используя уравнения движения Эйнштейна — Инфельда — Гофмана, а существенно позднее Леви-Чивита [238] подтвердил этот результат. Сегодня мы понимаем, что только неконсервативные теории гравитации могут предсказывать вековое ускорение центра масс и что измерения вековых изменений периода пульсара PSR 1913+16 могут служить проверкой таких теорий.

6. ГРАВИТАЦИЯ ВО ВСЕЛЕННОЙ: ВЛИЯНИЕ ГЛОБАЛЬНОЙ СТРУКТУРЫ НА ЛОКАЛЬНУЮ ФИЗИКУ

Поскольку гравитация проявляет себя через медленно спадающую с расстоянием силу чистого притяжения, можно ожидать, что глобальное распределение вещества во Вселенной влияет на локальную гравитационную физику. Действительно, с этой «махианской» точки зрения поистине удивительно, каким образом хоть какая-нибудь метрическая теория гравитации может избежать такого влияния. Однако общая теория относительности, очевидно, избегает этого влияния, вследствие чего о ней можно сказать, что она удовлетворяет сильному принципу эквивалентности (Сильн. ПЭ), который гласит:

1) *СПЭ и УГКС справедливы и* 2) *результат любого гравитационного или негравитационного эксперимента, проводимого в локальной свободно падающей системе отсчета, не зависит от того, где и когда во Вселенной этот эксперимент проводится, и не зависит от скорости этой системы отсчета* [ср. с ЭПЭ (п. 2.4), в котором речь идет только о негравитационных экспериментах].

Чтобы понять, почему общая теория относительности включает в себя Сильн. ПЭ, а большинство альтернативных теорий его не включают, необходимо рассмотреть, как решается локальная гравитационная проблема, например задача определения метрики пространства-времени в Солнечной системе, если локальная система погружена в реальную Вселенную, а не в асимптотически плоское пространство-время. Расчет метрики разбивается на две части: космологическое решение и «локальное» решение. С этой точки зрения Вселенная влияет на локальную гравитационную физику системы за счет «сшивки», т. е. за счет граничных условий (на границе «вдали» от вещества) для различных полей, создаваемых локальной системой. Из этого можно сделать ряд выводов.

1. *В теории, включающей в себя исключительно метрические поля, локальная гравитационная физика одинакова во всех асимптотических лоренцевых системах отсчета и не зависит от структуры или эволюции Вселенной, т. е. такая теория включает в себя Сильн. ПЭ.* Все это следует из свойств инвариантности метрики Минковского η (являющейся асимптотикой метрики g) — единственного поля, асимптотически связывающего локальную систему с Вселенной, а также из общей ковариантности, которая всегда позволяет

нам найти такую систему координат, в которой метрики на границе между Вселенной и локальной системой принимают вид метрики Минковского. Кроме того, были сделаны утверждения, что любая теория, включающая в себя Сильн. ПЭ, предсказывает полное отсутствие эффекта Нордтвердта (МТУ [1], § 20.6, 40.9), полное отсутствие вариаций в локально измеряемой гравитационной постоянной из-за окружающего вещества [12] и полное отсутствие дипольного гравитационного излучения [182, 183]. Общая теория относительности относится к классу именно таких теорий.

2. *В теории, содержащей наряду с метрическим полем некоторое скалярное поле* ϕ, *локальная физика одинакова во всех асимптотически лоренцевых системах отсчета, но может зависеть от структуры и эволюции Вселенной.* Эти выводы следуют из инвариантности как η, так и ϕ относительно преобразований Лоренца, но теперь ϕ может зависеть от структуры и эволюции Вселенной. Скалярно-тензорные теории относятся именно к этому типу, и, следовательно, Сильн. ПЭ в них может нарушаться.

3. *В теориях, включающих в себя кроме метрических векторные поля* **K**, *или тензорные поля* **C**, *или нединамические поля* t, η, *или их комбинацию, локальная гравитационная физика может зависеть от движения относительно выделенной универсальной системы покоя, а также от структуры и эволюции Вселенной.* Это обусловлено тем, что асимптотические величины дополнительных полей не обязательно инвариантны относительно преобразований Лоренца и принимают значения, определяемые космологической структурой. Это справедливо даже для плоской фоновой метрики η, как и в биметрических теориях (п. 3.2, б [105, 106]).

Из этого, по-видимому, следует, что любая теория, отличная от общей теории относительности, в которой вводятся дополнительные поля, динамические или нединамические, может предсказывать нарушения Сильн. ПЭ. Следовательно, влияние глобальной структуры на локальную гравитационную физику может приводить к наблюдаемым следствиям, которые включают в себя вековые вариации ньютоновской гравитационной постоянной (п. 6.1) и определяемые космологией значения ППН-параметров (п. 6.2). В последнем случае ограничения, налагаемые на ППН-параметры проверками в Солнечной системе, могут быть использованы для ограничения космологических моделей в теориях этого типа. Космологические модели, совместимые с локальной гравитационной физикой, могут после этого противоречить глобальным наблюдательным ограничениям, таким, как существование микроволнового фонового излучения.

6.1. ПОСТОЯНСТВО ГРАВИТАЦИОННОЙ ПОСТОЯННОЙ

Большинство теорий гравитации, в которых нарушается Сильн. ПЭ, предсказывают возможность изменения локально измеряемой

ньютоновской гравитационной постоянной с течением времени в ходе эволюции Вселенной. Для теорий, перечисленных в табл. 7, предсказания для $\dot{G}/G$ можно выразить через производные по времени от асимптотических динамических полей или через асимптотические значения параметров сшивки. Другие, более эвристические предположения об изменении гравитационной постоянной типа гипотез, высказанных Дираком, нельзя описать указанным способом. Дайсон [69] подробно обсудил такого рода предположения. Если бы константа G действительно менялась в ходе космологической эволюции, то темп ее вариации должен быть порядка темпа расширения Вселенной, т. е.

$$\dot{G}/G = \sigma H_0,$$

где H_0 — параметр хаббловского расширения, величина которого $H_0 \approx 55$ км·с$^{-1}$·Мпс$^{-1} \approx (2 \cdot 10^{10}$ лет$)^{-1}$, а σ — безразмерный параметр, значение которого зависит от рассматриваемой теории гравитации и от свойств конкретной космологической модели. Систематическое исследование значений параметра σ было проведено лишь для немногих теорий (обзор и библиографию для случая теории Бранса — Дикки см. в разд. 16.4 книги Вейнберга [122]). Однако можно наложить ряд наблюдательных ограничений на $\dot{G}/G$, воспользовавшись такими методами, как исследование эволюции скоплений галактик и Солнца, наблюдения лунных колеба-

Таблица 12

Наблюдения постоянства гравитационной постоянной

$\sigma = \dot{G}/G \cdot (2 \cdot 10^{10}$ лет)	Метод	Литература
$\lvert\sigma\rvert < 8 \cdot 10^{-1}$	Эволюция скоплений галактик	Дирборн и Шрам, 1974 [239], см. также Маршан и Мансфилд, 1977 [240]
$\lvert\sigma\rvert < 2$	Эволюция Солнца	Чин и Стосерс, 1976 [241]
$\lvert\sigma\rvert < 8 \cdot 10^{-1}$	Колебания Луны	Моррисон, 1973 [242]
$\sigma = -(7{,}2 \pm 3{,}6) \cdot 10^{-1}$	Колебания Луны	Ван Фландерн, 1975 [243, 244], 1976 [245]
$\lvert\sigma\rvert < 8$	Радиолокация планет	Шапиро и др., 1971 [246]
$\lvert\sigma\rvert < 2$	Радиолокация планет	Ризенберг и Шапиро, 1976 [247]
[a]	Лабораторные эксперименты	Брагинский и Гинзбург, 1974 [248]
		Риттер и др., 1976 [249]
		Брагинский и др., 1977 [172]

[a] Эксперименты, которые еще только предстоит провести.

ний, радиолокация планет и лабораторные эксперименты, которые еще только предстоит поставить. Краткий обзор современных наблюдений дан в табл. 12. Все имеющиеся результаты, за исключением измерений лунных колебаний, проведенных Ван Фландерном [245], совместимы с нулевыми космологическими вариациями константы G. Как отметили Ризенберг и Шапиро [247], поскольку ошибки радиолокационных наблюдений за изменением $\dot{G}/G$ падают как $T^{-3/2}$, где T — время, охватываемое наблюдениями, можно ожидать, что к 1985 г. точность такого метода достигнет $\Delta|\dot{G}/G|<10^{-11}$ год$^{-1}$. Андерсон, Кизи и др. [136], а также Вар и Бендер [159] показали, что радиолокационные наблюдния за «Викингом» и за аппаратом, выведенным на орбиту вокруг Меркурия, за два года полета могли бы дать точность $\Delta|\dot{G}'/G|\sim10^{-12}$ год$^{-1}$.

6.2. КОСМОЛОГИЧЕСКИЕ ПРОВЕРКИ ГРАВИТАЦИОННЫХ ТЕОРИЙ

Поскольку общая теория относительности удовлетворяет Сильн. ПЭ, наблюдательная космология включает в себя лишь измерение чисто космологических параметров, таких, как параметр хаббловского расширения H_0, параметр плотности $\Omega_0=8\pi G\rho_0/3H_0^2$, где ρ_0— сегодняшняя средняя плотность Вселенной, параметр ускорения q_0 и т. д. (более подробные определения космологических параметров см. в МТУ [1], § 29). В случае однородных и изотропных космологических моделей с равной нулю космологической постоянной для того, чтобы однозначно задать космологическую модель, достаточно двух из вышеуказанных параметров, скажем H_0 и Ω_0. (Чтобы определить поведение модели в прошлом, в любом случае необходимо знать уравнение состояния вещества или, что эквивалентно, плотность излучения.) Однако в случае теорий, нарушающих Сильн. ПЭ, влияние глобальной структуры на локальную гравитационную физику приводит к появлению новых параметров, которые необходимо определять, например это отношение измеряемой гравитационной постоянной G к «чистой» (затравочной) константе взаимодействия G_*, которую дает теория; это темп изменения или значения некоторых или всех ППН-параметров. Например, в скалярно-тензорных теориях, чтобы однозначно задать космологическую модель, необходимо знать H_0, Ω_0, ϕ_0, $\dot{\phi}_0$ и функцию $\omega(\phi)$, тогда как в биметрической теории Розена (с плоской фоновой метрикой) модель однозначно определяется величинами H_0, Ω_0, $(c_0)_0$, $(c_1)_0$ (п. 3.2, б) или H_0, Ω_0, $(\dot{G}/G)_0$ и ППН-параметром $(\alpha_2)_0$, где индекс 0 указывает на то, что значения величины берутся в настоящий момент [250]. Когда наблюдаемые значения этих параметров используют в качестве граничных условий, то может возникать вопрос, согласуются ли космологические модели в этих теориях с такими наблюдениями, как существование и изотропия микроволнового фонового излучения, распространенности гелия и дейтерия,

наблюдаемые значения или ограничения на параметр q_0 и т. д. В теории Бранса — Дикки, например, такие модели, которые согласуются с наблюдениями, построить можно, правда, чем больше значение ω (требуемое экспериментами в Солнечной системе, разд. 3), тем более близки эти модели по своим свойствам к моделям в общей теории относительности. Однако в биметрической теории Розена и в теории расслоения Ни космологические модели, согласующиеся с современными значениями космологических и локальных параметров, предсказывают в некоторых случаях «отскок» в прошлом, а не «большой взрыв». Кейвз [250] показал, что такой отскок возникает при характерных плотностях и температурах более низких, чем это необходимо для ионизации водорода. Следовательно, в подобных моделях очень трудно достичь такого сжатия, чтобы возникло наблюдаемое микроволновое фоновое излучение, и невозможно получить наблюдаемую распространенность гелия с помощью первичного нуклеосинтеза.

Если Вселенная анизотропна, то теории, нарушающие Сильн. ПЭ, могут предсказывать анизотропные ППН-параметры [251]. Нордтведт [219, 251], воспользовавшись измерениями земных приливов и наблюдениями смещения периастра двойного пульсара, установил пределы на некоторые из таких анизотропий; эти пределы лежат в интервале от 10^{-4} до 10^{-9}. Отметим, что недавние наблюдения микроволнового фона указывают на изотропность Вселенной, по крайней мере, с точностью $3 \cdot 10^{-4}$ [161].

Сама Вселенная, возможно, станет ареной захватывающего, очень точного сопоставления гравитационной теории и эксперимента.

7. ЗАКЛЮЧЕНИЕ

Свое обсуждение сопоставления гравитационной теории с экспериментом мы начали вблизи от нашего дома, в лаборатории, в которой проверялись основы гравитационной теории. Но наше поле зрения быстро расширялось, охватывая Солнечную систему, гравитационное излучение, приходящее из космоса, двойной пульсар и весь космос. Это расширение поля зрения экспериментальной гравитации идет параллельно с непрерывным расширением горизонтов астрономии: по мере прогресса техники методы измерений в астрономии способны проникать все дальше, становясь при этом все более точными. Проникая в глубины, мы вновь и вновь обнаруживаем ту фундаментальную роль, которую играет гравитация в нашей Вселенной, и со все большей остротой нам необходима правильная теория гравитации. Тот же самый технический прогресс, благодаря которому ширятся астрономические горизонты, дает все новые средства для сопоставления гравитационной теории с экспериментом в совершенно новых областях, все выше становится уровень точности. Как свидетельствуют статьи в этом сборнике, общей

теории относительности придерживаются большинство теоретиков-гравитационистов и астрофизиков. Но этого недостаточно. Общую теорию относительности следует сопоставлять с каждым новым экспериментальным фактом и, только основываясь на эксперименте, можно судить об ее истинности или ложности. До сих пор она выдерживала любое такое сопоставление, но новые сопоставления, в новых областях, уже не за горами. Выдержит ли общая теория относительности все испытания? Для одних это вопрос чисто умозрительный, для других это вопрос веры, а некоторые твердо знают, что выдержит. Но независимо от теоретической позиции ни у кого не может вызывать сомнений, что гравитация — наиболее давно известное и во многих смыслах самое фундаментальное из всех взаимодействий — заслуживает более надежного эмпирического обоснования.

Мне приятно поблагодарить за плодотворные обсуждения Боба Вагонера, Марка Хогена, Джона Андерсона, Карла Кейвза и Джо Тейлора, а также Викторию Ла Брие за перепечатку рукописи.

ЛИТЕРАТУРА[1])

1. *Misner C. W., Thorne K. S., Wheeler J. A.*, Gravitation, Freeman, San Francisko, 1973. [Имеется перевод: *Мизнер Ч. В., Торн К. С., Уилер Дж. А.*, Гравитация.— М.: Мир, 1977.]
2. *Will C. M.*, The theoretical tools of experimental gravitation, In: Experimental Gravitation: Proceedings of Course 56 of the International School of Physics «Enrico Fermi», ed. B. Bertotti, pp. 1—110. Academic Press, London, New York, 1974.
3. *Richard J. P.*, Tests of theories of gravitiy in the solar system, In: General Relativity and Gravitation, eds. G. Shaviv and J. Rosen, pp. 169—188, Wiley, New York, 1975.
4. *Brill D. R.*, Observational contacts of General Relativity. In: Relativity, Astrophysics, and Cosmology, ed. W. Israel, pp. 127—52. Reidel, Dordrecht, 1973.
5. *Nordtvedt K., Jr.*, Gravitation theory: Empirical status from solar-system experiments. Science, **178**, 1157—64 (1972).
6. *Will C. M.*, Einstein on the firing line. Physics Today, **25**, No. 10, 23—9 (1972).
7. *Will C. M.*, Gravitation theory. Scientific American, **231**, No. 5, 25—33 (1974).
8. *Dicke R. H.*, Experimental relativity. In: Relativity, Groups and Topology, eds. C. DeWitt and B. DeWitt, pp. 165—313. Gordon and Breach, New York, 1964.
9. *Dicke R. H.*, Remarks on the observational basis of General Relativity. In: Gravitation and Relativity, eds. H.-Y. Chiu and W. F. Hoffman, pp. 1—16, Benjamin, New York. 1964.
10. *Bertotti B., Brill D. R., Krotkov R.*, Experiments on gravitation. In: Gravitation: An Introduction to Current Research, ed. L. Witten, pp. 1—48. Wiley, New York, 1962.
11. *Thorne K. S., Will C. M.*, Theoretical frameworks for testing relativistic gravity. I. Foundations. Astrophys. J., **163**, 595—610 (1971).
12. *Will C. M., Nordtvedt K., Jr.*, Conservation laws and preferred frames in rela-

---

[1]) Литература, отмеченная звездочкой, добавлена при переводе.— *Прим. перев.*

tivistic gravity. I. Preferred-frame theories and an extended PPN formalism. Astrophys. J., **177**, 757—74 (1972).
13. *Thorne K. S., Will C. M. Ni W.-T.*, Theoretical frameworks for testing relativistic gravity—A review. In: Proceedings of the Conference on Experimental Tests of Gravitation Theories, ed. K. W. Davies, pp. 10—31, NASA-JPL Technical Memorandum 33—499, 1971.
14. *Thorne K. S., Lee D. L., Lightman A. P.*, Foundations for a theory of gravitation theories. Phys. Rev., **D7**, 3563—78 (1973).
15. *Will C. M.*, Gravitational redshift measurements as tests of non-metric theories of gravity. Phys. Rev., **D10**, 2330—7 (1974).
16. *Mansouri R., Sexl R. U.*, A test theory of Special Relativity. I. Simultaneity and clock synchronization. Gen. Relativ. Grav., **8**, 497—513 (1977).
17. *Mansouri R., Sexl R. U.*, A text theory of Special Relativity. II. First order tests. Gen Relativ. Grav., **8**, 515—24 (1977).
18. *Hughes V. W., Robinson H. G., Beltran-Lopez V.*, Upper limit for the anistropy of inertial mass from nuclear resonance-experiments. Phys. Rev. Lett., **4**, 342—4 (1960).
19. *Drever R. W. P.*, A search for anisotropy of inertial mass using a free precession technique. Philos. Mag., **6**, 683—7, (1961).
20. *Peebles P. J. E., Dicke R. H.*, Significance of spatial isotropy. Phys. Rev., **127**, 629—31 (1962).
21. *Peebles P. J. E.*, The Eötvös experiment, spatial isotropy, and generally covariant field theories of gravity. Ann. Phys. (N. Y.), 20, 240—60 (1962).
22. *Nordtvedt K., Jr., Haugan M. P.*, в печати.
23. *Wagoner R. V.*, Scalar-tensor theory and gravitational waves. Phys. Rev., **D1**, 3209—16 (1970).
24. *Fujii Y.*, Dilaton and possible non-Newtonian gravity. Nature (Phys. Sci.), **234**, 5—7 (1971).
25. *Fujii Y.*, Scale invariance and gravity of hadrons. Ann. Phys. (N. Y.), **69**, 494—521.
26. *O'Hanlon J.*, Intermediate-range gravity: A generally covariant model. Phys. Rev. Lett., **29**, 137—8 (1972).
27. *Long D. R.*, Experimental examination of the gravitational inverse square law. Nature, **260**, 417—18 (1976).
28. *Newman R., Pellam J., Schultz J., Spero R.*, Experimental test of the gravitational inverse square law at laboratory distances. In: Proceedings of the 8th International Conference on General Relativity and Gravitation (не опубликовано), 268, University of Waterloo, Canada, 1977.
29. *Paik H. J., Mapoles E., Fairbank W. M.*, частное сообщение, 1977.
30. *Mikkelson D. R., Newman M. J.*, Constraints on the gravitational constant at large distances. Phys. Rev., **D16**, 919—26 (1977).
31. *Koester L.*, Verification of the equivalence of gravitational and inertial mass for the neutron. Phys. Rev., **D14**, 907—9 (1976).
32. *Fairbank W. M., Witteborn F. C., Madey J. M. J., Lockhart J. M.*, Experiments to determine the force of gravity of single electrons and positrons. In: Experimental Gravitation: Proceedings of Course 56 of the International School of Physics «Enrico Fermi», ed. B. Bertotti, pp. 310—30. Academic Press, London, New York, 1974.
33. *Newton I.*, Philosophiae Naturalis Principia Mathematica, London, 1687.
34. *Bessel F. W.*, Poggendorff's Ann., **25**, 401 (1832).
35. *Eötvös R. V., Pekár V., Fekete E.*, Beitrage zum Gesetze der Proportionalität von Trägheit und Gravität. Ann. Phys. (Leipzig), **68**, 11—66 (1922).
36. *Potter H. H.*, Some experiments on the proportionality of mass and weight. Proc. R. Soc. Lond., **104**, 588—610 (1923).
37. *Renner J.*, Mat. Termeszettud. Ert., **53**, 542 (1935).
38. *Roll P. G., Krotkov R., Dicke R. H.*, The equivalence of inertial and passive gravitational mass. Ann. Phys. (N. Y.), **26**, 442—517 (1964).
39. *Брагинский В. Б., Панов В. И.*, Проверка эквивалентности инертной и гравитационной массы, ЖЭТФ, **61**, 873—879 (1971).

40. *Worden P. W., Jr.*, A cryogenic test of the Equivalence Principle. Ph. D. Thesis, Stanford University, 1976.
41. *Worden P. W., Jr., Everitt C. W. F.*, The gyroscope experiment. III. Tests of the equivalence of gravitational and inertial mass based on cryogenic techniques. In: Experimental Gravitation: Proceedings of Course 56 of the International School of Physics «Enrico Fermi», ed. B. Bertotti, pp. 381—402, Academic Press, London, New York, 1974.
42. *Pound R. V., Rebka G. A., Jr.*, Apparent weight of photons. Phys. Rev. Lett., **4**, 337—41 (1960).
43. *Pound R. V., Snider J. L.*, Effect of gravity on gamma radiation. Phys. Rev., 140, B788—803 (1965).
44. *Brault J. W.*, The gravitational redshift in the solar spectrum. Ph. D. Thesis, Princeton University, 1962.
45. *Jenkins R. E.*, A satellite observation of the relativistic Doppler shift. Astron. J., **74**, 960—3 (1969).
46. *Snider J. L.*, New measurement of the solar gravitational redshift. Phys. Rev. Lett., **28**, 853—6 (1972).
47. *Hafele J. C., Keating R. E.*, Around-the-world atomic clocks: Predicted relativistic time gains. Science, **177**, 166—8, (1972).
48. *Hafele J. C., Keating R. E.*, Around-the-world atomic clocks: Observed relativistic time gains. Science, **177**, 168—70 (1972).
49. *Alley C. O., Cutler L. S., et al.*, Atomic clock measurements of the General Relativistic time differences produced by aircraft flights using both direct and laser-pulse remote time comparison (готовится к печати).
50. *Vessot R. F. C., Levine M. W.*, A preliminary report of the gravitational redshift rocket-probe experiment. In: Proceedings of the 2nd Frequency Standards and Metrology Symposium, ed. H. Hellwig, pp. 659—88. National Bureau of Standards, Boulder, Colorado, 1976.
51. *Turneaure J. P., Will C. M.*, A null gravitational redshift experiment. Bull. Am. Phys. Soc., **20**, 1488 (1975).
52. *Will C. M.*, Clocks and experimental gravitation: A null gravitational redshift experiment, laboratory tests of post-Newtonian gravity, and gravity-wave detection by spacecraft tracking. Metrologia, **13**, 95—8 (1977).
53. *Nordtvedt K., Jr.*, A study of one-and two-way Doppler tracking of a clock on an arrow toward the Sun. In: Proceedings of the International Meeting on Experimental Gravitation, ed. B. Bertotti, pp. 247—56, Accademia Nazionale dei Lincei, Rome 1977.
54. *Vessot R. F. C.*, Lectures on frequency stability and clocks and on the gravitational redshift experiment. In: Experimental Gravitation: Proceedings of Course 56 of the International School of Physics «Enrico Fermi», ed. B. Bertotti, pp. 111—62, Academic Press, London, New York, 1974.
55. *Stein S. R.*, The superconducting-cavity stabilized oscillator and an experiment to detect time variation of the fundamental constants. Ph. D. Thesis, Stanford University, 1974.
56. *Stein S. R., Turneaure J. P.*, Superconducting-cavity stabilized oscillators with improved frequency stability. IEEE Proc., **63**, 1249—50 (1975).
57. *McGuigan D. F., Douglass D. H.*, Clocks based upon high mechanical Q single crystals. In: Proceedings of the 31st Annual Frequency Control Symposium (в печати).
58. *Jaffe J., Vessot R. F. C.*, Feasibility of a second-order gravitational redshift experiment. Phys. Rev., **D14**, 3294—300 (1976).
59. *Schiff L. I.*, On experimental tests of the General Theory of Relativity. Am. J. Phys., **28**, 340—3 (1960).
60. *Lightman A. P., Lee D. L.*, Restricted proof that the Weak Equivalence Principle implies the Einstein Equivalence Principle. Phys. Rev., **D8**, 364—76 (1973).
61. *Nordtvedt K., Jr.*, Gravitational and inertial mass of bodies of interacting electrical charges. Int. J. Theor. Phys., **3**, 133—9 (1970).
62. *Haugan M. P., Will C. M.*, Principles of equivalence, Eötvös experiments,

and gravitational redshift experiments: The free fall of electromagnetic systems to post-post-Coulombian order. Phys. Rev., **D15**, 2711—20 (1977).
63. *Ni W. T.*, Equivalence principles and electromagnetis. Phys. Rev. Lett., **38**, 301—4 (1977).
64. *Nordtvedt K., Jr.*, Quantitative relationship between clock gravitational redshift violations and non-universality of free-fall rates in non-metric theories of gravity. Phys. Rev., **D11**, 245—7 (1975).
65. *Haugan M. P.*, Energy conservation and the principle of equivalence, Ann. Phys. (N. Y.) (в печати).
66. *Haugan M. P., Will C. M.*, Weak interactions and Eötvös experiments. Phys. Rev. Lett., **37**, 1—4 (1976).
67. *Chapman P. K., Hanson A. J.*, An Eötvös experiment in Earth orbit. In: Proceedings of the Conference on Experimental Tests of Gravitation Theories, ed. R. W. Davies, pp. 228—35, NASA-JPL Technical Memorandum 33—499, 1971.
68. *Chiu H.-Y., Hoffman W. F.*, Introduction. In: Gravitation and Relativity, eds. H. Y. Chiu and W. F. Hoffman, pp. XIII—XXXV. Benjamin, New York, 1964.
69. *Dyson F. J.*, The fundamental constants and their time variation. In: Aspects of Quantum Theory, eds. A. Salam and E. P. Wigner, pp. 213—36. Cambridge University Press, 1972,
70. *Wolfe A. M., Brown R. L., Roberts M. S.*, Limits on the variation of fundamental atomic quantities over cosmic time scales. Phys. Rev. Lett., **37**, 179—81 (1976).
71. *Shliakhter A. I.*, Direct test of the constancy of fundamental nuclear constants. Nature, 264, 340 (1976).
72. *Solheim J.-E., Barnes T. G., III, Smith H. J.*, Obervational evidence against a time variation in Planck's constant. Astrophys. J., **209**, 330—4, (1976).
73. *Baum W. A., Florentin-Nielsen R.*, Cosmological evidence against time variation of the fundamental atomic constants. Astrophys. J., **209**, 319—29 (1976).
74. *Turneaure J. P., Stein S. R.*, An experimental limit on the time variation of the fine structure constant. In: Atomic Masses and Fundamental Costants, vol. 5, eds. J. H. Sanders and A. H. Wapstra, pp. 636—42. Plenum, New York, 1976.
75. *Pagel B. E. J.*, On the limits to past variability of the proton-electron mass ratio set by quasar absorption lines. Mon. Not. R. Astron. Soc., **179**, 81P—85P (1977).
76. *Davies P. C. W.*, Time variation of the coupling constants. J. Phys., **A5**, 1296—1304 (1972).
77. *Will C. M.*, Relativistic gravity in the solar system. III. Experimental disproof of a class of linear theories of gravitation. Astrophys. J., **185**, 31—42 (1973).
78. *Lee D. L., Lightman A. P., Ni W.-T.*, Conservation Laws and variational principles in metric theories of gravity. Phys. Rev., **D10**, 1685—1700 (1974).
79. *Eddington A. S.*, The Mathematical Theory of Relativity. Cambridge University Press, 1922.
80. *Robertson H. P.*, Relativity and cosmology. In: Space Age Astronomy, eds. A. J. Deutsch and W. B. Klemperer, pp. 228—35, Academic Press, London, New York, 1962.
81. *Schiff L. I.*, Comparison of theory and observation in General Relativity. In: Relativity Theory and Astrophysics. I. Relativity and Cosmology, ed. J. Ehlers, pp. 105—16. American Mathematical Society, Providence, 1967.
82. *Will C. M.*, Relativistic gravity in the solar system. II. Anisotropy in the Newtonian gravitational constant. Astrophys. J., **169**, 141—55 (1971).
83. *Nordtvedt K., Jr., Will C. M.*, Conservation laws and preferred frames in relativistic gravity. II. Experimental evidence to rule out preferred-frame theories of gravity. Astrophys., J., **177**, 775—92 (1972).
84. *Will C. M.*, Theoretical frameworks for testing relativistic gravity. III Con-

servation laws, Lorentz invariance, and values of the PPN parameters. Astrophys. J., **169**, 125—40 (1971).
85. *Deser S., Laurent B.*, Linearity and parametrization of gravitational effects. Astron. Astrophys., **25**, 327—8 (1973).
86. *Duff M. J.*, On the significance of perihelion shift calculations. Gen. Relativ. Grav., **5**, 441—52 (1974).
87. *Shiff L. I.*, Motion of a gyroscope according to Einstein's theory of gravitation. Proc. Nat. Acad. Sci. USA, **46**, 871—82 (1960).
88. *Nordtvedt K., Jr.*, Equivalence principle for massive bodies. II. Theory. Phys. Rev., **169**, 1017—25 (1968).
89. *Will C. M.*, Theoretical frameworks for testing relativistic gravity II. Parametrized post-Newtonian hydrodynamics and the Nordtvedt effect. Astrophys. J., **163**, 611—28 (1971).
90. *Baierlein R.*, Testing General Relativity with laser ranging to the Moon. Phys. Rev., **162**, 1275—88 (1967).
91. *Will C. M.*, Active mass in relativistic gravity: Theoretical interpretation of the Kreuzer experiment. Astrophys. J., **204**, 224—34 (1976).
92. *Nordtvedt K., Jr.*, Testing relativity with laser ranging to the Moon. Phys. Rev., **170**, 1186—7 (1968).
93. *Anderson J. L.*, Principles of Relativity Physics. Academic Press, London, New York, 1967.
94. *Ni W.-T.*, Theoretical frameworks for testing relativistic gravity. IV. A compendium of metric theories of gravity and their post-Newtonian limits. Astrophis. J., **176**, 769—96 (1972).
95. *Bergmann P. G.*, Comments on the Scalar-Tensor theory. Int. J. Theor. Phys., **1**, 25—36 (1968).
96. *Nordtvedt K., Jr.*, Post-Newtonian metric for a general class of Scalar-Tensor gravitational theories and observational consequences. Astrophys. J., **161**, 1059—67 (1970).
97. *Bekenstein J. D.*, Are particle rest masses variable? Theory and constraints from solar system experiments. Phys. Rev., **D15**, 1458—68 (1977).
98. *Brans C., Dicke R. H.*, Mach's principle and a relativistic theory of gravitation. Phys. Rev., **124**, 925—935 (1961).
99. *Hellings R. W., Nordtvedt K. J.*, Vector-metric theory of gravity. Phys. Rev., **D7**, 3593—3602 (1973).
100. *Rosen N.*, A bi-metric theory of gravitation., Gen. Relativ. Grav., **4**, 435—47 (1973).
101. *Rosen N.*, A theory of gravitation. Ann. Phys. (N. Y.), **84**, 455—73, (1974).
102. *Posen N.*, Bimetric gravitation and cosmology. Astrophys. J., **211**, 357—60 (1977).
103. *Rosen N., Rosen J.*, The maximum mass of a cold neutron star. Astrophys. J., **202**, 782—7 (1975).
104. *Rosen N.*, Bimetric theory on a cosmological basis., Gen. Relativ. Grav., **9**, 339—45 (1978).
105. *Lee D. L., Caves C. M., Ni W. T., Will C. M.*, Theoretical frameworks for testing relativistic gravity. V. Post-Newtonian limit of Rosen's theory. Astrophys. J., **206**, 555—8 (1976).
106. *Lightman A. P., Lee D. L.*, New two-metric theory of gravity with prior geometry. Phys. Rev., **D8**, 3293—302 (1973).
107. *Rastall P.*, A theory of gravity. Can. J. Phys., **54**, 66—75 (1976).
108. *Rastall P.*, A note on a theory of gravity. Can. J. Phys., **55**, 38—42 (1977).
109. *Rastall P.*, Conservation laws and gravitational radiation. Can. J. Phys., **55**, 1342—8 (1977).
110. *Ni W.-T.*, A new theory of gravity. Phys. Rev., **D7**, 2880—3 (1963).
111. *Whitehead A. N.*, The Principle of Relativity. Cambridge University Press, 1922.
112. *Shapiro I. I.*, New method for the detection of light deflection by solar gravity. Science, **157**, 806—8 (1967).

113. *Ward W. R.*, General Relativistic light deflection for the complete celestial sphere. Astrophys. J., **162**, 345—8 (1970).
114. *Shapiro I. I.*, Fourth test of General Relativity. Phys. Rev. Lett., **13**, 789—91 (1964).
115. *Merat P., Pecker J. C., Vigier J. P., Yourgrau W.*, Observed deflection of light by the Sun as a function of solar distance. Astron. Astrophys., **32**, 471—5 (1974).
116. Texas Mauritanian Eclipse Team. Gravitational deflection of light: solar eclipse of 30 June 1973. I. Description of procedures and final results. Astron. J., **81**, 452—4, (1976).
117. *Jones B. F.*, Gravitational deflection of light: solar eclipse of 30 June 1973. II Plate reductions. Astron. J., **81**, 455—63, (1974).
118. *Hill H. A.*, Light deflection. In: Proceedings of the Conference on Experimental Tests of Gravitation Theories, ed. R. W. Davies, pp. 89—91. NASA-JPL Technical Memorandum 33—499 (1971).
119. *Muhleman D. O., Ekers R. D., Fomalont E. B.*, Radio interferometric test of the general relativistic light bending near the Sun. Phys. Rev. Lett., **24**, 1377—80 (1970).
120. *Seielstad G. A., Sramek R. A., Weiler K. W.*, Measurement of the deflection of 9.602-GHz radiation from 3C279 in the solar gravitational field. Phys. Rev. Lett., **24**, 1373—6 (1970).
121. *Hill J. M.*, A measurement of the gravitational deflection of radio waves by the Sun. Mon. Not. R. Astron. Soc., **153**, 7P-11P (1971).
122. *Weinberg S.*, Gravitation and Cosmology. Wiley, New York, 1972. [Имеется перевод: *Вейнберг С.* Гравитация и относительность.— М.: Мир, 1976.]
123. *Sramek R. A.*, A measurement of the gravitational deflection of microwave radiation near the Sun. 1970 October. Astrophys. J., **167**, L55—60 (1971).
124. *Sramek R. A.*, The gravitational deflection of radio waves. In: Experimental Gravitation: Proceedings of Course 56 of the International School of Physics «Enrico Fermi», ed. B. Bertotti, pp. 529—42. Academic Press, London, New York, 1974.
125. *Riley J. M.*, A measurement of the gravitational deflection of radio waves by the Sun during October 1972. Mon. Not. R. Astron. Soc., **161**, 11P—14P (1973).
126. *Weiler K. W., Ekers R. D., Raimond E., Wellington K. J.*, A measurement of solar gravitational microwave deflection with the Westerbork synthesis telescope. Astron. Astrophys., **30**, 241—8 (1974).
127. *Counselman C. C.*, III, *Kent S. M., et al.*, Solar gravitational deflection of radio waves measured by very-long-baseline interferometry. Phys. Rev. Lett., **33**, 1621—3 (1974).
128. *Weiler K. W., Ekers R. D., Raimond E., Wellington K. J.*, Dual-frequency measurement of the solar gravitational microwave deflection. Phys. Rev. Lett., **35**, 134—7 (1975).
129. *Fomalont E. B., Sramek R. A.*, A confirmation of Einstein's General Theory of Relativity by measuring the bending of microwave radiation in the gravitational field of the Sun. Astrophys. J., **199**, 749—55 (1975).
130. *Fomalont E. B., Sramek R. A.*, Measurement of the solar gravitational deflection of radio waves in agreement with General Relativity. Phys. Rev. Lett., **36**, 1475—8, (1976).
131. *Fomalont E. B., Sramek R. A.*, The deflection of radio waves by the Sun. Comment Astrophys., 7, 19—33 (1977).
132. *Muhleman D. O., Reichley P.*, Effects of General Relavity on planetary radar distance measurements. JPL Space Programs Summary 4 No. 37—39, 239—41 (1964).
133. *Shapiro I. I.*, Fourth test of General Relativity: preliminary results. Phys. Rev. Lett., 20, 1265—9 (1968).
134. *Shapiro I. I., Ash M. E., Ingalls R. P., et al.*, Fourth test of General Relativity: new radar result. Phys. Rev. Lett., 26, 1132—5 (1971).
135. *Anderson J. D., Esposito P. B., Martin W., Thornton C. L., Muhleman D. O.*,

Experimental test of General Relativity using time-delay data from Mariner 6 and Mariner 7., Astrophys. J., **200**, 221—33 (1975).
136. *Anderson J. D., Keesey M. S. W., et al.*, Tests of General Relativity using astrometric and radiometric observations of the planets. In: Proceedings of the Third International Space Relativity Symposium (27th Congress, International Astronautical Federation), в печати.
137. *Reasenberg, R. D., Shapiro I. I.*, Solar-system tests of General Relativity. In: Proceedings of the International Meeting on Experimental Gravitation, ed. B. Bertotti, pp. 143—60. Accademia Nazionale dei Lincei, Roma, 1977.
138. *Shapiro I. I., Reasenberg R. D., et al.*, The Viking relativity experiment. J. Geophys. Res., **82**, 4329—34 (1977).
139. *Anderson J. D.*, Lectures on physical and technical problems posed by precision radio tracking. In: Experimental Gravitation: Proceedings of Course 56 of the International School of Physics «Enrico Fermi», ed. B. Bertotti, pp. 163—99. Academic Pres, London, New York, 1974.
140. *Shapiro I. I.*, Testing General Relativity: progress, problems and prospects. Gen Relativ. Grav., **3**, 135—48 (1972).
141. *Nordtvedt K., Jr.*, Equivalence principle for massive bodies. I. Phenomenology. Phys. Rev., **169**, 1014 (1968). II. Theory. Phys. Rev. **169**, 1017 (1968).
142. *Nordtvedt K., Jr.*, Equivalence principle for massive bodies including rotational energy and radiation pressure. Phys. Rev., **180**, 1293—8 (1969).
143. *Nordtvedt K., Jr.*, Equivalence principle for massive bodies. IV. Planetary orbits and modified Eötvös-type experiments. Phys. Rev., **D3**, 1683—9 (1971).
144. *Dicke R. H.*, Gravitation and the Universe. American Philosophical Society, Philadelphia, 1969.
145. *Nordtvedt K., Jr.*, Post-Newtonian gravitational effects in lunar laser ranging. Phys. Rev., **D7**, 2347—56 (1973).
146. *Will C. M.*, Relativistic gravity in the solar system. I. Effect of an anisotropic gravitational mass on the Earth-Moon distance. Astrophys. J., **165**, 409—12 (1971).
147. *Bender P. L., Currie D. G., Dicke R. H., et al.*, The lunar laser ranging experiment. Science, **182**, 229—38 (1973).
148. *Williams J. G., Dicke R. H., Bender P. L., et al.*, New test of the equivalence principle from lunar laser ranging. Phys. Rev. Lett., **36**, 551—4 (1976).
149. *Shapiro I. I., Counselman C. C., III King R. W.*, Verification of the principle of equivalence for massive bodies. Phys. Rev. Lett., **36**, 555—8 (1976).
150. *Nordtvedt K., Jr.*, Solar-system Eötvös experiments. Icarus, **12**, 91—100 (1970).
151. *Nordtvedt K., Jr.*, Tests of the equivalence principle and gravitation theory using solar system bodies. In: Proceedings of the Conference on Experimental Tests of Gravitation Theories, ed. R. W. Davies, pp. 32—7, NASA—JPL Technical Memorandum 33—499, 1971.
152. *Morrison L. V., Ward C. G.*, An analysis of the transits of Mercury: 1667—1973. Mon. Not. R. Astron. Soc., **173**, 183—206 (1975).
153. *Anderson J. D., Colombo G., Friedman L. D., Lau E. L.*, An arrow to the Sun. In: Proceedings of the International Meeting on Experimental Gravitation, ed. B. Bertotti, pp. 393—422. Accademia Nazionale dei Lincei, Rome, 1977.
154. *Shapiro I. I., Pettengill G. H., et al.*, Mercury's perihelion advance: determination by radar. Phys. Rev. Lett., **28**, 1594—7 (1972).
155. *Dicke R. H., Goldenberg H. M.*, The oblateness of the Sun. Astrophys. J. Suppl., **27**, 131—82 (1974).
156. *Dicke R. H.*, The oblateness of the Sun and relativity. Science, **184**, 419—29 (1974).
157. *Hill H. A., Clayton P. D., Patz D.* L., et al., Solar oblateness, excess brightness and relativity. Phys. Rev. Lett., **33**, 1497—500 (1974).
158. *Hill H. A., Stebbins R. T.*, The intrinsic visual oblateness of the Sun. Astrophys. J., **200**, 471—83 (1975).
159. *Wahr J. M., Bender P. L.*, Determination of PPN parameters from Earth—Mercury distance measurements. Preprint, 1976.

160. *Rose R. D., Parker H. M., Lowry R. A., Kuhlthau A. R., Beams J. W.*, Determination of the gravitational constant G. Phys. Rev. Lett., **23**, 655—8 (1969).
161. *Smoot G. F., Gorenstein M. V., Muller R. A.*, Detection of anisotropy in the cosmic blackbody radiation. Phys. Rev. Lett., **39**, 898—901 (1977).
162. *Rochester M. G., Smylie D. E.*, On changes in the trace of the Earth's inertia tensor. J. Geophys. Res., **79**, 4948—51 (1974).
163. *Warburton R. J., Goodkind J. M.*, Search for evidence of a preferred reference frame. Astrophys. J., **208**, 881—6 (1976).
164. *Schiff L. I.*, Possible new test of General Relativity Teory. Phys. Rev. Lett., **4**, 215—17 (1960).
165. *Everitt C. W. F.*, The gyroscope experiment. I. General description and analysis of gyroscope performance. In: Experimental Gravitation: Proceedings of Course 56 of the International School of Physics «Enrico Fermi», ed. B. Bertotti, pp. 331—60. Academic Press, London, New York, 1974.
166. *Lipa J. A., Fairbank W. M., Everitt C. W. F.*, The gyroscopy experiment. II. Development of the London-Moment gyroscope and of cryogenic technology for space. In: Experimental Gravitation: Proceedings of Course 56 of the International School of Physics «Enrico Fermi», ed. B. Bertotti, pp. 361—80. Academic Press London, New York, 1974.
167. *Wilkins D. C.*, General equation for the precession of a gyroscope. Ann. Phys. (N. Y.), **61**, 277—93 (1970).
168. *Van Patten R. A., Everitt C. W. F.*, Possible experiment with two counter-orbiting drag-free satellites to obtain a new test of Einstein's General Theory of Relativity and improved measurements in geodesy. Phys. Rev. Lett., **36**, 629—32 (1976).
169. *Kreuzer L. B.*, Experimental measurement of the equivalence of active and passive gravitational mass. Phys. Rev., **169**, 1007—12 (1968).
170. *Gilvarry J. J., Muller P. M.*, Possible variation of the gravitational constant over the elements. Phys. Rev. Lett., **28**, 1665—8 (1972).
171. *Morrison D., Hill H. A.* Current uncertainty in the ratio of active-to-passive gravitational mass. Phys. Rev., **D8**, 2731—3 (1973).
172. *Braginsky V. B., Caves C. M., Thorne K. S.*, Laboratory experiments to test relativistic gravity. Phys. Rev., **D15**, 2047—68 (1977).
173. *Eardley D. M., Lee D. L., Lightman A. P., Wagoner R. V., Will C. M.*, Gravitational-wave observations as a tool for testing relativistic gravity. Phys. Rev. Lett., **30**, 884—6 (1973).
174. *Eardley D. M., Lee D. L., Lightman A. P.*, Gravitational-wave observations as a tool for testing relativistic gravity. Phys. Rev., **D10**, 3308—21 (1973).
175. *Paik H. J.*, Response of a disk antenna to scalar and tensor gravitational waves. Phys. Rev., **D15**, 409—15 (1977).
176. *Wagoner R. V., Paik H. J.*, Multi-mode detection of gravitational waves by a sphere. In: Proceedings of the International Meeting on Experimental Gravitation, ed. B. Bertotti, pp. 257—66. Accademia Nazionale dei Lincei, Rome, 1977.
177. *Thorne K. S.*, The generation of gravitational waves. V. Multipole-moment formalisms. Preprint. Cornell University, CRSR 663, 1977.
178. *Peters P. C., Mathews J.*, Gravitational radiation from point masses in Keplerian orbit. Phys. Rev., **131**, 435—40 (1963).
179. *Ehlers J., Rosenblum A., Goldberg J. J., Havas P.*, Astrophys. J., **208**, L77—81 (1976).
180. *Wagoner R. V.*, Test for the existence of gravitational radiation. Astrophys. J., **196**, L63—5 (1975).
181. *Faulkner J.*, Ultrashort-period binaries, gravitational radiation, and mass transfer. I. The standard model, with applications to WZ Sagittae and Z Camelopardalis. Astrophys. J., **170**, L99—104 (1971).
182. *Eardley D. M.*, Observable effects of a scalar gravitational field in a binary pulsar. Astrophys. J., **196**, L59—62 (1975).
183. *Will C. M., Eardley D. M.*, Dipole gravitational radiation in Rosen's theory

of gravity: Observable effects in the binary system PSR 1913+16. Astrophys. J., **212**, L91—4 (1977).
184. *Will C. M.*, Gravitational radiation from binary systems in alternative metric theories of gravitation: Dipole radiation and the binary pulsar. Astrophys. J., **214**, 826—39 (1977).
185. *Barker B. M.*, General Scalar-Tensor theory of gravity with constant G. Astrophys J., **219**, 5—11 (1978).
186. *Epstein R., Wagoner R. V.*, Post-Newtonian generation of gravitational waves. Astrophys. J., **197**, 717—23 (1975).
187. *Hulse R. A., Taylor J. H.*, Discovery of a pulsar in a binary system. Astrophys. J., **195**, L51—3 (1975).
188. *Taylor J. H., Hulse R. A., et al.*, Further observations of the binary pulsar PSR 1913+16. Astrophys J., **206**, L53—8 (1976).
189. *Blandford R., Teukolsky S. A.*, Arrival-time analysis for a pulsar in a binary system. Astrophys. J., **205**, 580—91, (1976).
190. *Taylor J. H.*, Discovery of a pulsar in a binary system. Ann. N. Y. Acad. Sci., **262**, 490—2 (1975).
191. *Masters A. R., Roberts D. H.*, On the nature of the binary system containing the pulsar PSR 1913+16. Astrophys J., **195**, L107—11 (1975).
192. *Webbink R. F.*, PSR 1913+16: Endpoints of speculation. A critical discussion of possible companions and progenitors. Astron. Astrophys., **41**, 1—8 (1975).
193. *Davidsen A., Margon B., et al.*, Optical and X-ray observations of the PSR 1913+16 field. Astrophys. J., **200**, L19—21 (1975).
194. *Kristian J., Clardy K. D., Westphal J. A.*, Upper limits for the visible counterpart of the Hulse—Taylor binary pulsar. Astrophys. J., **206**, L143—4 (1976).
195. *Roberts D. H., Masters A. R., Arnett W. D.*, Determining the stellar masses in the binary system containing the pulsar PSR 1913+16: Is the companion a helium main-sequence star? Astrophys. J., **203**, 196—201 (1976).
196. *Flannery B. P., van den Heuvel E. P. J.*, On the origin of the binary pulsar PSR 1913+16. Astron. Astrophys., **39**, 61—7 (1975).
197. *Van Horn H. M., Sofia S., et al.*, Binary pulsar PSR 1913+16: model for its origin. Science, **188**, 930—3 (1975).
198. *van den Berg S.*, The binary pulsar 1913+16. Astrophys. Lett. **16**, **75** (1975).
199. *Wheeler J. C.*, The binary pulsar: preexplosion evolution. Astrophys. J., **205**, 578—9 (1976).
200. *Smarr L. L., Blandford R.*, The binary pulsar: Physical processes, possible companions, and evolutionary histories. Astrophys. J., **207**, 574—88 (1976).
201. *Бисноватый-Коган Г. С., Комберг Б. В.*, Радио-пульсар в двойной системе — старый объект со слабым магнитным полем: возможная эволюционная схема его образования. Письма в Астрон. журн., **2**, 338—342 (1976).
202. *Озерной Л. М., Шишов В. И.*, Верхний предел электронной концентрации вблизи пульсара PSR 1913+16 в двойной системе. Письма в Астрон. журн., **1**, 21—24 (1975).
203. *Balbus S. A., Brecher K.*, Tidal friction in the binary pulsar system PSR 1913+16. Astrophys. J., **203**, 202—5 (1976).
204. *Shapiro S. L., Terzian Y.*, Galactic rotation and the binary pulsar. Astron. Astrophys., **52**, 115—18 (1976).
205. *Hjellming R. M., Gibson D. M.*, An interferometric search for the Hulse—Taylor binary pulsar. Astrophys. J., 199, L165—6 (1975).
206. *Chanan G., Middleditch J., Nelson J.*, An upper limit on optical pulsations from PSR 1913+16. Astrophys. J., **199**, L167—8 (1975).
207. *Bernacca P. L., Ciatti F., et al.*, Search for an optical counterpart of the binary pulsar PSR 1913+16. Astron. Astrophys., **40**, 327—9, (1975).
208. *Nather R. E., Robinson E. L., Van Citters G. W., Hemenway P. D.*, An upper limit to optical pulses from the binary pulsar PSR 1913+16. Astrophys J., **211**, L125—7 (1977).
209. *Van Citters G. W., Rybski P. M.*, Area photometry in the region of the pulsar 1913+16. Astrophys J., **214**, 233—4 (1977).

210. *Wheeler J. C.*, Timing effects in pulsed binary systems. Astrophys. J., **196**, L67—70 (1975).
211. *Blandford R.*, *Teukolsky S. A.*, On the measurement of the mass of PSR 1913+16. Astrophys J., **198**, L27—9 (1975).
212. *Epstein R.*, The binary pulsar: post-Newtonian timing effects. Astrophys J., **216**, 92—100 (1977).
213. *Озерной Л. М.*, *Рейнгард М.*, О природе компаньона радиопульсара PSR 1913+16 в составе двойной системы. Письма в Астрон. журн., **1**, 16—20 (1975).
214. *Damour T.*, *Ruffini R.*, Sur certaines vérifications nouvelles de la Relativite Générale rendues possibles par la découverte d'un pulsar membre d'un système binaire. C. R. Acad. Sci. (Paris), **279**, A971—3 (1974).
215. *Brecher K.*, Some implications of period changes in the first binary radio pulsar Astrophys J., **195**, L113—5 (1975).
216. *Брумберг В. А.*, *Зельдович Я. Б.*, *Новиков И. Д.*, *Шакура Н. И.*, Определение масс компонентов и наклона двойной системы, содержащей пульсар, по релятивистским эффектам, Письма в Астрон. журн., **1**, 5—9 (1975).
217. *Esposito L. W.*, *Harrison E. R.*, Properties of the Hulse — Taylor binary pulsar system. Astrophys. J., **196**, L1-2 (1975).
218. *Demianski M.*, *Shakura N. I.*, A secular relativistic change in the period of a binary pulsar. Nature, **263**, 665—6 (1976).
219. *Nordtvedt K.*, *Jr.*, Anisotropic gravity and the binary pulsar PSR 1913+16. Astrophys. J., **202**, 248—9 (1975).
220. *Will C. M.*, A test of post-Newtonian conservation laws in the binary system PSR 1913+16. Astrophys. J., **206**, 861—7 (1976).
221. *Will C. M.*, Periastron shifts in the binary system PSR 1913+16. Theoretical interpretation. Astrophys. J., **196**, L3—5 (1975).
222. *Barker B. M.*, *O'Connell R. F.*, Relativistic effects in the binary pulsar PSR 1913+16. Astrophys. J., **199**, L25—6 (1975).
223. *Зельдович Я. Б.*, *Шакура Н. И.*, Релятивистский эффект неравномерности вращения пульсара на эллиптической орбите, Письма в Астрон. журн., **1**, 19—22 (1975).
224. *Hari Dass N. D.*, *Radhakrishnan V.*, The new binary pulsar and the observation of gravitational spin precession. Astrophys. Lett., **16**, 135—9 (1975).
225. *Rudolph E.*, Relativistic observable effects in the binary pulsar PSR 1913+16. In: Proceedings of the International School of Physics «Enrico Fermi», 1976. Academic Press, London, New York, 1977.
226. *Wagoner R. V.*, A new test of General Relativity. Gen Relativ. Grav., **7**, 333—7 (1976).
227. *Caporaso G.*, *Brecher K.*, Neutron-star mass limit in the bimetric theory of gravitation. Phys. Rev., **D15**, 3536—42 (1977).
228. *Mikkelson D. R.*, Very massive neutron stars in Ni's theory of gravity. Astrophys. J., **217**, 248—51 (1977).
229. *Rastall P.*, The maximum mass of a neutron star. Astrophys. J., **213**, 234—8 (1977).
230. *Rosen N.*, Is there gravitational radiation? Nuovo Cimento Lett., **19**, 249—50 (1977).
231. *Press W. H.*, *Wiita P. J.*, *Smarr L. L.*, Mechanism for inducing synchronous rotation and small eccentricity in close binary systems. Astrophys. J., **202**, L135—7 (1975).
232. *Shapiro S. L.*, *Teukolsky S. A.*, On the maximum gravitational redshift of white dwarfs. Astrophys J., **203**, 697—700 (1976).
233. *Wagoner R. V.*, *Malone R. C.*, Post-Newtonian neutron stars., Astrophys. J., **189**, L75—8 (1974).
234. *Levi-Civita T.*, Astronomical consequences of the relativistic two-body problem. Am. J. Math., **59**, 225—34 (1937).
235. *Eddington A. S.*, *Clark G. L.*, The problem of *n* bodies in General Relativity theory. Proc. R. Soc. Lond., **166**, 465—75 (1938).

236. *de Sitter W.*, On Einstein's theory of gravitation and its astronomical consequences. Mon. Not. R. Astron. Soc., **77**, 155—84 (1916).
237. *Robertson H. P.*, The two-body problem in General Relativity. Ann. Math., **39**, 101—4 (1938).
238. *Levi-Civita T.*, The n-Body Problem in General Relativity. Reidel, Dordrecht, 1964.
239. *Dearborn D. S.*, *Schramm D. N.*, Limits on variation of *G* from clusters of galaxies. Nature, **247**, 441—3 (1974).
240. *Marchant A.*, *Mansfield V.*, Evolution of dynamical systems with time-varying gravity. Nature, **270**, 699—700 (1977).
241. *Chin C.-W.*, *Stothers R.*, Limit on the secular change of the gravitational constant based on studies of solar evolution. Phys. Rev. Lett., **36**, 833—5 (1976).
242. *Morrison L. V.*, Rotation of the Earth from AD 1663—1972 and the constancy of *G*. Nature, 241, 519—20 (1973).
243. *Van Flandern T. C.*, A determination of the rate of change of *G*. Mon. Not. R. Astron. Soc., **170**, 333—42 (1975).
244. *Van Flandern T. C.*, Recent evidence for variations in the value of *G*. Ann. N. Y. Acad. Sci., **262**, 494—5 (1975).
245. *Van Flandern T. C.*, Is gravity getting weaker? Scientific American, **234**, no. 2, 44—52 (1976).
246. *Shapiro I. I.*, *Smith W. B.*, *Ash M. E.*, *Ingalls R. P.*, *Pettengill G. H.*, Gravitational constant: experimental bound on its time variation. Phys. Rev. Lett., **26**, 27—30 (1971).
247. *Reasenberg R. D.*, *Shapiro I. I.*, Bound on the secular variation of the gravitational interaction. In: Atomic Masses and Fundamental Constants, vol. 5, eds. J. H. Sanders and A. H. Wapstra, pp. 643—9. Plenum, New York, 1976.
248. *Брагинский В. Б.*, *Гинзбург В. Л.* О возможности измерения зависимости гравитационной постоянной от времени, ДАН, **216**, 300—302 (1974).
249. *Ritter R. C.*, *Beams J. W.*, *Lowry R.A.*, A laboratory experiment to measure the time variation of Newton's gravitational constant. In: Atomic Masses and Fundamental Constants, vol. 5, eds., J. H. Sanders and A. H. Wapstra, pp. 629—35. Plenum, New York, 1976.
250. *Caves C. M.*, Cosmological observations as tests of relativistic gravity: Rosen's bimetric theory. In: Proceedings of the 8th International Conference on General Relativity and Gravitation, 104. University of Waterloo, Canada, 1977.
251. *Nordtvedt K., Jr.*, Anisotropic parametrized post-Newtonian gravitational metric field. Phys. Rev., **D14**, 1511—17 (1976).
252*. *Reasenberg D. D.*, *Shapiro I. I.*, *et al.*, Viking relativity experiment: Verification of signal retardiation by solar gravity. Astrophys., J., **234**, L219—221 (1979).
253*. *Thorne K. S.*, Gravitational-wave research: Current status and future prospects. Rev. of Modern Phys., **52**, 285—297 (1980).
254*. *Weisberg J. M.*, *Taylor J.*, Gravitational radiation from an orbiting pulsar. General Rel. and Grav., **13**, 1—6 (1981).

II. ПРОБЛЕМА НАЧАЛЬНЫХ ДАННЫХ И ДИНАМИЧЕСКАЯ ФОРМУЛИРОВКА ОБЩЕЙ ТЕОРИИ ОТНОСИТЕЛЬНОСТИ

А. Фишер, Дж. Марсден[1])

Эта статья посвящена обсуждению ряда взаимосвязей между задачей Коши, каноническим формализмом, устойчивостью линеаризации и пространством гравитационных степеней свободы. В последнее десятилетие наблюдается возрождение интереса к этим вопросам, поскольку в ходе развития математических методов и подходов между ними стали выясняться тесные связи. В настоящее время литература, касающаяся этих разделов общей теории относительности, представляет собой быстро расширяющуюся область.

В этой статье мы стремимся дать представление о текущем состоянии предмета с нашей точки зрения. Для установления связей между упомянутыми выше вопросами будут использованы развитые нами геометрические методы. При этом основными инструментами будут служить нелинейный функциональный анализ, формализм сопряженных величин для гамильтоновых полевых теорий и бесконечномерная симплектическая геометрия. Как мы увидим, этот математический аппарат и рассматриваемые вопросы естественным образом связаны между собой. Для более полного ознакомления с нынешним состоянием предмета читатель отсылается к работам [7, 35, 53, 103, 104, 121—124, 144, 150].

В разд. 1 развивается гамильтонов формализм для динамики общей теории относительности, называемый обычно формализмом АДМ (Арновитта — Дезера — Мизнера). При этом используются инвариантные понятия и метод сопряженных величин, разработанный авторами. Будет показано, как можно представить динамическую систему Эйнштейна в следующей явной и компактной форме:

$$\text{Уравнение эволюции} \begin{pmatrix} \dfrac{\partial g}{\partial \lambda} \\ \dfrac{\partial \pi}{\partial \lambda} \end{pmatrix} = J \circ [D\Phi(g, \pi)]^{*} \cdot \begin{pmatrix} N \\ X \end{pmatrix},$$

$$\text{Уравнение связи } \Phi(g, \pi) = 0.$$

Такой вид уравнений полезен для изучения устойчивости линеаризации и пространства гравитационных степеней свободы. Мы наметим схему, позволяющую распространить формализм сопря-

---

[1]) *A. E. Fischer, J. E. Marsden.*, Division of Natural Sciences, University of California, Santa Cruz, USA.

женных величин на все теории, в которых связь с гравитацией минимальна.

Формализм сопряженных величин естественным образом ведет к изучению многообразия связи (разд. 2); основным результатом этого раздела является ответ на вопрос, какие точки принадлежат многообразию (регулярны), а какие являются точками бифуркации (сингулярны). Мы покажем также (используя формализм сопряженных величин), что подмногообразие связи находится в инволюции относительно динамических уравнений. Уравнения, используемые для получения этого результата, эквивалентны каноническим коммутационным соотношениям Дирака.

На основе этого динамического формализма в разд. 3 и 4 обсуждаются существование, единственность и устойчивость решения задачи Коши. В разд. 3 резюмируется общая теория гиперболической задачи с начальными данными, которая потребуется нам для теории относительности. Мы излагаем абстрактный подход, из которого как частные случаи следуют результаты, касающиеся существования и единственности для симметричных гиперболических систем первого порядка, для гиперболических систем второго порядка и для комбинации этих систем. Приводимые нами теоремы являются наиболее сильными из известных результатов в отношении дифференцируемости. Соответственно их применение в разд. 5 приводит к наиболее сильным результатам в отношении существования и единственности для задач Коши для уравнений поля в пустом пространстве (теоремы 23 и 27). Попутно будут сделаны замечания о том, как применить эту абстрактную теорию к полям, взаимодействующим с гравитацией.

Несмотря на значительный прогресс в изучении проблемы начальных данных, остается нерешенной главная проблема связи динамических сингулярностей (несуществование решений уравнений эволюции «для всех времен») с сингулярностями в смысле Хокинга — Пенроуза.

В разд. 5, комбинируя результаты разд. 2 и 4, мы получаем условия, при которых первый порядок теории возмущений имеет силу, и показываем, что при наличии вектора Киллинга ряд теории возмущений должен быть перестроен для согласования с существованием этого вектора. Здесь же приведены необходимые условия второго порядка для того, чтобы возмущение было интегрируемо. Этими результатами мы обязаны совместной работе с В. Монкри.

Наконец, в разд. 6 обсуждается исключение калибровочных условий с помощью общей процедуры редукции для гамильтоновых систем. Применяя этот общий метод, мы показываем далее, что пространство гравитационных степеней свободы является, вообще говоря, бесконечномерным симплектическим многообразием. Таким образом, множество геометрий пустого пространства есть в общем случае бесконечномерное гравитационное фазовое пространство без сингулярностей. Наш общий формализм может быть при-

менен также и к полям, взаимодействующим с гравитацией минимально; без особого труда можно показать, что пространство степеней свободы для этих полей и гравитации в общем случае также является симплектическим многообразием.

Дальнейшие сведения по этим вопросам читатель может найти в работах [2—4, 54, 92—95, 108, 114, 142, 145—149, 173].

Авторы благодарны Дж. Армсу, И. Шоке-Брюа, К. Кухаржу, В. Монкри, Р. Палэ, Р. Саксу и А. Таубу за полезные советы, а также С. Хокингу и В. Израэлю за любезное предложение участвовать в этом сборнике.

1. КАНОНИЧЕСКИЙ ФОРМАЛИЗМ

Напомним сначала четырехмерный лагранжев формализм в классической теории полей, взаимодействующих с гравитацией. Вслед за этим будет изложен динамический, или «3+1»-подход.

Используются следующие обозначения: V_4— гладкое четырехмерное многообразие; под термином «многообразие» подразумевается связное, ориентируемое, паракомпактное, хаусдорфово многообразие; TV_4— его касательное расслоение. Положим также, что

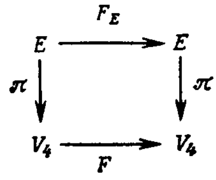
$L(V_4)$ = множество всех гладких лоренцевых метрик с сигнатурой $(-+++)$;

$S_2(V_4)$ = множество всех гладких симметричных тензорных полей 2-го ранга на V_4.

Пусть теперь E есть векторное расслоение над V_4 с проекцией $\pi : E \to V_4$, а его C^∞-сечения обозначаются $C^\infty(E)$. Мы часто будем иметь дело с $E = T^r_s(V_4)$, расслоением тензоров с r контравариантными и s ковариантными индексами. Однако важно понимать, что это не означает принципиального ограничения лишь теорией тензорных полей: тогда из рассмотрения исключались бы такие важные полевые теории, как теория Янга — Миллса; см., например, [3, 105]. Строго говоря, поля Янга — Миллса требуют использования аффинного расслоения (расслоения связностей на главном расслоении над V_4), но для этого не нужно сколь-нибудь существенно переделывать рассматриваемый нами формализм.

Если введены какие-либо координаты, мы будем записывать компоненты $\varphi \in C^\infty(E)$ как φ_A, где A — собирательный индекс для всего набора индексов.

Обозначим через $\mathscr{D}(V_4)$ (сохраняющие ориентацию) диффеоморфизмы многообразия V_4. Для «естественных расслоений» любое $F \in \mathscr{D}(V_4)$ расширяется функторно до диффеоморфизма расслоения $F_E\colon E \to E$, накрывающего F, т. е. диаграмма



коммутативна, и $(F\circ G)_E = F_E \circ G_E$. При этом F_E есть обычное преобразование тензоров, если $E = T^r_s(V_4)$. Тогда *увлечение назад* F определено на сечениях E и действует следующим образом:

$$F^*\colon C^\infty(E) \to C^\infty(E);\ \varphi \to F_E^{-1}\circ\varphi\circ F = F^*\varphi,$$

а его обращение, *увлечение вперед*, определяется равенством $F_*\varphi = F_E\circ\varphi\circ F^{-1}$. Для расслоений, связанных с полями Янга — Миллса, в дополнение к понятиям увлечения вперед и назад, которые порождаются $\mathscr{D}(V_4)$, имеется еще бесконечномерная калибровочная группа.

Заметим, что E может быть суммой Уитни $E_1 \oplus E_2 \oplus \ldots E_k$ для k типов полей, так что наш формализм пригоден и для взаимодействующих полей.

Пусть Ω — расслоение плотностей (т. е. 4-форм) над V_4; E^* — *дуальное расслоение* над V_4, слоем которого в точке x является $E^*_x \otimes \Omega_x$, где E^*_x — векторное пространство, дуальное к E_x. Таким образом, E^* есть расслоение векторных плотностей над V_4. Например, если $E = T^r_s(V_4)$, то $E^* = T^r_s(V_4)\otimes\Omega$ будет расслоением тензорных плотностей типа $\binom{s}{r}$.

Мы имеем естественное L_2-спаривание между $C^\infty(E)$ и $C^\infty(E^*)$, задаваемое как

$$(\varphi,\ \psi\otimes d\mu)_{L_2} = \int_{V_4} \psi(\varphi)\, d\mu,$$

где $\psi(\varphi)d\mu = \psi(\varphi)\otimes d\mu$ и подразумевается интегрируемость $\psi(\varphi)$ по $d\mu$. Мы будем говорить о $C^\infty(E^*)$ как о сечении, *естественным образом L_2-дуальном* к $C^\infty(E)$.

Пусть E и F — два расслоения над V_4 и $A\colon C^\infty(E) \to C^\infty(F)$ есть линейный оператор. *Оператор, естественным образом сопряженный* к A, определяется как $A^*\colon C^\infty(F^*) \to C^\infty(E^*)$, $(A^*(\tilde\psi), \varphi)_{L_2} = (\tilde\psi, A\varphi)_{L_2}$ для $\tilde\psi \in C^\infty(F^*)$, $\varphi \in C^\infty(E)$. Здесь, конечно, молчаливо предполагается, что A^* существует.

Если A — дифференциальный оператор, то A^* вычисляется, как обычно, интегрированием по частям, что дает сопряженный дифференциальный оператор. Вообще говоря, A^* можно понимать в смысле неограниченного оператора [109].

Для расслоения E над V_4 с дуальным расслоением E^* будем считать, что E в свою очередь является дуальным к E^*, так что $(E^*)^* = E$. Таким образом, если $E = T^r_s(V_4)$, то $E^* = T^s_r(V_4)\otimes\Omega$ и $(E^*)^* = T^r_s(V_4)$. При этом условии, если $A\colon C^\infty(E) \to C^\infty(F^*)$, то $A^*\colon C^\infty(F) \to C^\infty(E^*)$.

Рассмотрим теперь лагранжеву плотность в теории полей, взаимодействующих с гравитацией.

$$\mathscr{L}\colon L(V_4)\times C^\infty(E) \to C^\infty_d(V_4),$$

где $C^\infty_d(V_4) = \Omega$ есть расслоение скалярных плотностей над V_4. Запишем $\mathscr{L}(g, \varphi) = \mathscr{L}_{\text{грав}}(g) + \mathscr{L}_{\text{полей}}(g, \varphi)$, причем $\mathscr{L}_{\text{грав}}(g) = (^1/_{16}\pi)\, R(g) d\mu(g)$, где $R(g)$ — скалярная кривизна метрики g

и $d\mu(g) = (-\det g_{\alpha\beta})^{1/2}\, dx^0 \wedge dx^1 \wedge dx^2 \wedge dx^3$ — элемент объема, определяемый $g \in L(V_4)$; позднее мы будем обозначать такое g как ${}^{(4)}g$.

Если потребовать, чтобы интеграл действия

$$S(g, \varphi) = \int_{\mathcal{D}} [\mathcal{L}_{\text{грав}}(g) + \mathcal{L}_{\text{полей}}(g, \varphi)]\, d\mu(g)$$

был стационарным для любой ограниченной открытой области $\mathcal{D} \subset V_4$ с гладкой границей $\partial\mathcal{D}$ и для любой вариации h метрики g и вариации ψ полей φ, обращающихся в нуль на этой границе, то получим

$$0 = \int_{\mathcal{D}} [\mathrm{D}\mathcal{L}_{\text{грав}}(g) \cdot h + \mathrm{D}_g \mathcal{L}_{\text{полей}}(g, \varphi) \cdot h + \mathrm{D}_\varphi \mathcal{L}_{\text{полей}}(g \cdot \varphi) \cdot \psi]\, d\mu(g)$$

для всех h и ψ, исчезающих на $\partial\mathcal{D}$, где через D обозначена производная Фреше, а D_g, D_φ — частные производные соответственно по g и φ. Отметим, что вариации h и ψ берутся соответственно из $S_2(V_4)$, пространства симметричных дважды ковариантных тензорных полей на V_4 и из $C^\infty(E)$.

В терминах естественным образом сопряженных операторов это условие превращается в уравнения Эйлера — Лагранжа:

$$[\mathrm{D}\mathcal{L}_{\text{грав}}(g)]^* \cdot 1 + [\mathrm{D}_g \mathcal{L}_{\text{полей}}(g, \varphi)]^* \cdot 1 = 0$$

и

$$[\mathrm{D}_\varphi \mathcal{L}_{\text{полей}}(g, \varphi)]^* \cdot 1 = 0,$$

где 1 есть постоянная функция 1 в пространстве функций с действительными значениями, которое является дуальным к пространству плотностей $C_d^\infty(V_4)$. Эти уравнения эквивалентны обычному способу записи уравнений Эйлера — Лагранжа (если предположить, что $\mathcal{L}_{\text{полей}}$ зависит от k-струи g, φ):

$$\frac{\delta\mathcal{L}_{\text{грав}}}{\delta g} - \frac{\delta\mathcal{L}_{\text{полей}}}{\delta g} = 0,$$

$$\frac{\delta\mathcal{L}_{\text{полей}}}{\delta\varphi} = 0.$$

Тогда, согласно [132], имеем

$$\mathrm{D}R(g) \cdot h = \Delta\, \mathrm{tr}\, h + \delta\delta h - h \cdot \mathrm{Ric}(g),$$

где Δ = оператор Лапласа — Бельтрами на скалярах; $\Delta f = -f_{;\alpha}{}^{;\alpha}$,
tr = след; $\mathrm{tr}\, h = h^\alpha{}_\alpha$,
$\delta h = -\mathrm{div}\, h = -h_\alpha{}^\beta{}_{;\beta}$,
$\delta\delta h$ = двойная дивергенция $= h^{\alpha\beta}{}_{;\alpha;\beta}$,
$\mathrm{Ric}(g)$ = тензор Риччи для $g = R_{\alpha\beta}$
и $\mathrm{D}[d\mu(g)] \cdot h = (1/2)(\mathrm{tr}\, h)\, d\mu(g)$.

Таким образом,

$$\mathrm{D}\mathscr{L}_{\text{грав}}(g)\cdot h=\frac{1}{16\pi}[\Delta\operatorname{tr}h+\delta\delta h-\operatorname{Ein}(g)\cdot h]\,d\mu(g),$$

где $\operatorname{Ein}(g)=\operatorname{Ric}(g)-{}^1/_2 g\,R(g)$ — тензор Эйнштейна для g (т. е. $G_{\alpha\beta}=R_{\alpha\beta}-{}^1/_2 g_{\alpha\beta}R$). Поскольку интеграл от $\Delta\operatorname{tr}h+\delta\delta h$ обращается в нуль для вариаций h, исчезающих на $\partial\mathscr{D}$, отсюда следует, что

$$[\mathrm{D}\mathscr{L}_{\text{грав}}(g)]^{*}\cdot 1=-\frac{1}{16\pi}[\operatorname{Ein}(g)]^{\#}d\mu(g),$$

где $\#$ означает, что индексы подняты с помощью g.

Положим (см. [104], § 3.3)

$$\mathscr{T}(g,\varphi)=2\frac{\delta\mathscr{L}_{\text{полей}}}{\delta g}=2[\mathrm{D}_g\mathscr{L}_{\text{полей}}(g,\varphi)]^{*}\cdot 1\in S^2_d(V_4),$$

где $S^2_d(V_4)=S^2(V_4)\otimes\Omega$ означает пространство дважды контравариантных симметричных тензорных плотностей в V_4.

Пусть $T(g,\varphi)=\mathscr{T}(g,\varphi)^{*}$ — дуальный тензор в $S_2(V_4)$, индуцированный метрикой g. Таким образом, $\mathscr{T}=T^{\#}d\mu(g)$ и T является обычным симметричным тензором энергии-импульса, ассоциированным с $\mathscr{L}_{\text{полей}}(g,\varphi)$.

Полевые уравнения тогда примут вид

$$\operatorname{Ein}(g)=8\pi T(g,\varphi)\quad(\text{т. е. } G_{\mu\nu}=8\pi T_{\mu\nu})$$

$$\text{и}\quad \delta\mathscr{L}_{\text{полей}}/\delta\varphi=0.$$

Если мы хотим иметь достаточно хорошую теорию полей, необходимо наложить жесткие ограничения на возможный выбор $\mathscr{L}_{\text{полей}}$. Например, если наша теория тензорная и $\mathscr{L}_{\text{полей}}$ зависит от производных g, скажем от ковариантных производных $\nabla\varphi$ полей φ, то T, вообще говоря, зависит от вторых производных g и вторых производных φ. Подобным же образом уравнения для полей будут зависеть как от вторых производных метрики, так и от вторых производных полей. В такой ситуации может и не быть хорошо определенной системы гиперболических уравнений [121—123]. Поэтому обычно предполагается *минимальное взаимодействие* (полей с гравитацией), т. е. $\mathscr{L}_{\text{полей}}$ зависит лишь от значений g в точке. Для скалярного поля, электромагнитного поля и поля Янга — Миллса (для последнего $\mathscr{L}_{\text{полей}}=({}^1/_{16}\pi)F\cdot F\cdot d\mu(g)$ и $F=dA+[A,A]$ есть кривизна поля связности A) трудностей не возникает ввиду минимальности взаимодействий этих систем. Для минимальных тензорных теорий поля можно классифицировать содержащиеся в них *естественные* дифференциальные операторы согласно [151, 156, 172].

Перейдем теперь к динамической формулировке Дирака — АДМ. Для изложения этого предмета воспользуемся современной симплектической геометрией и неявным вариантом дираковской теории связей; см. [1]. Поскольку гравитация играет у нас выделенную роль, вначале рассмотрим ее. Затем мы сделаем несколько

замечаний относительно случая полей, взаимодействующих с гравитацией.

Как и раньше, пусть V_4— четырехмерное многообразие с лоренцевой метрикой ${}^{(4)}g$, ориентируемое и ориентируемое по времени. Мы пишем ${}^{(4)}g$, чтобы не смешивать ее с римановой метрикой g, которая вводится ниже. Пусть M — *компактное* ориентируемое трехмерное многообразие[1]) и пусть $i: M \to V_4$ есть вложение M, такое, что вложенное многообразие $i(M)=\Sigma$ пространственноподобно, т. е. сужение $i^*({}^{(4)}g)=g$ есть риманова метрика на M. Пусть $C^\infty_{\text{простр}}(M;\ V,\ {}^{(4)}g)$ означает множество всех таких пространственноподобных вложений. Как и в работе [76], оно является гладким многообразием. Пусть k означает вторую фундаментальную форму вложения, определенную в точке $m \in M$ для $X,\ Y \in T_m M$ обычной формулой

$$k_m(X,\ Y) = -{}^{(4)}g \circ i(m) \cdot ((T_m i \cdot Y),\ {}^{(4)}\nabla_{(T_m i \cdot X)}\,{}^{(4)}Z_\Sigma \circ i(m)),$$

где ${}^{(4)}Z_\Sigma \circ i(m)$ — направленная в будущее времениподобная нормаль к Σ в точке $i(m)$. Таким образом, $k_{ij}=Z_{i;j}$ (где точка с запятой означает ковариантное дифференцирование относительно метрики ${}^{(4)}g$; ковариантное дифференцирование относительно g обозначается вертикальной чертой).

Пусть $\pi=\pi' \otimes d\mu(g)$ — дважды контрвариантная тензорная плотность, тензорная часть π которой определяется как $\pi' = [(\operatorname{tr} k)g - k]^\#$, где $\#$ означает контрвариантную форму ковариантного тензора с индексами, поднятыми с помощью g; аналогично $\flat$ означает ковариантную форму контравариантного тензора. В гамильтоновой формулировке Арновитта — Дезера — Мизнера (АДМ) k играет роль переменной скорости, а π — ее канонического импульса. Отметим, что у нас $\pi=\pi^{\text{АДМ}} d^3x$. При обсуждении пространства гравитационных степеней свободы в разд. 6 полезно будет знать, что если $(V_4,{}^{(4)}g)$ глобально гиперболично и есть поверхность Коши, диффеоморфная M, то любое пространственноподобное вложение M в V_4 также является поверхностью Коши [24, 104].

Предположим теперь, что дана кривая в $C^\infty_{\text{простр}}(M;\ V_4,\ {}^{(4)}g)$, т. е. кривая i пространственноподобных вложений M в $(V_4,\ {}^{(4)}g)$. Производная по параметру λ (λ-производная) этой кривой задает однопараметрическое семейство векторных полей ${}^{(4)}X_{\Sigma_\lambda}$ на вложенных гиперповерхностях следующим уравнением:

$$\frac{di_\lambda}{d\lambda} = {}^{(4)}X_{i_\lambda} = {}^{(4)}X_{\Sigma_\lambda} \circ i_\lambda : M \to TV_4$$

(рис. 1). Нормальные и касательные проекции ${}^{(4)}X_{\Sigma_\lambda}$ задают кривую функций $N_\lambda = {}^{(4)}X_\perp : M \to \mathbb{R}$ и векторных полей ${}^{(4)}X_\parallel =$

[1]) Гамильтонов формализм для некомпактного случая имеет заметные отличия (см. [54, 161]), но теория существования и единственности, обсуждаемая в разд. 3, 4, справедлива и для этого случая.

$=X_\lambda : M \to TM$ на M уравнением

$$^{(4)}X_{\Sigma_\lambda} \circ i_\lambda(m) = {}^{(4)}X_\perp(\lambda, m)\, {}^{(4)}Z_{\Sigma_\lambda} \circ i_\lambda(m) + T_m i_\lambda \cdot {}^{(4)}X_\parallel(\lambda, m),$$

где $^{(4)}Z_{\Sigma_\lambda}$ — направленная в будущее единичная времениподобная нормаль к Σ_λ. Пусть $N_\lambda > 0$; тогда отображение

$$F: I \times M \to V_4;\ (\lambda, m) \to i_\lambda(m)$$

есть диффеоморфизм многообразия $I \times M$ на трубчатую окрестность многообразия $i_0(M) = \Sigma_0$, если интервал $I = (-\beta, \beta)$ выбран достаточно малым. В этом случае мы назовем и кривую i_λ, и вложенные гиперповерхности $\Sigma_\lambda = i_\lambda(M)$ *разбиением* многообразия V_4.

По терминологии Арновитта — Дезера — Мизнера [7] и Уилера [174] функции N_λ и векторные поля X_λ суть *функции длительности* и *векторные поля сдвига*.

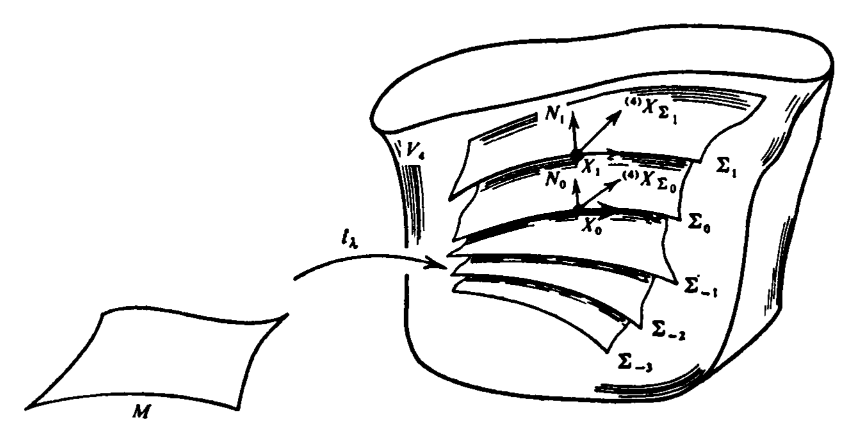


Рис. 1. Пространственноподобные вложения M в V_4 и разложение генератора $^{(4)}X_{\Sigma_\lambda}$ на нормальную и касательную составляющие.

Используя $F: I \times M \to V_4$ как координатную систему для трубчатой окрестности на гиперповерхности Σ_0 в V_4, координаты (x^i), $i = 1, 2, 3$ на M и $(x^\alpha) = (\lambda, x^i)$, $\alpha = 0, 1, 2, 3$ как координаты на $I \times M$, получим суженную метрику $F^{*(4)}g$ в виде

$$(F^{*(4)}g)_{\alpha\beta}\, dx^\alpha dx^\beta = -(N^2 - X_i X^i)\, d\lambda^2 + 2X_i\, dx^i\, d\lambda + g_{ij}\, dx^i\, dx^j,$$

где $g_{ij} = (g_\lambda)_{ij}$ и $g_\lambda = i_\lambda^{*(4)}g$.

Пусть k_λ — кривая вторых фундаментальных форм вложенных гиперповерхностей $\Sigma_\lambda = i_\lambda(M)$ и пусть π_λ — связанные с ними канонические импульсы. Основные геометродинамические уравнения, введенные в работах [7, 32, 72, 63, 130], содержатся в следующей теореме.

Теорема 1

Пусть на V_4 выполняются вакуумные уравнения Эйнштейна $\mathrm{Ein}(^{(4)}g)=0$. *Тогда для каждого однопараметрического семейства пространственноподобных вложений* $\{i_\lambda\}$ *многообразия* V_4 *индуцированные на* Σ_λ *метрики* g_λ *и импульсы* π_λ *удовлетворяют следующим уравнениям:*

$$\text{Уравнения эволюции} \begin{cases} \dfrac{\partial g}{\partial \lambda} = 2N\left[(\pi') - \frac{1}{2} g(\operatorname{tr}\pi')\right] + L_X g, \\ \dfrac{\partial \pi}{\partial \lambda} = -2N\left[\pi' \times \pi' - \frac{1}{2}(\operatorname{tr}\pi')\pi'\right] d\mu(g) + \end{cases}$$

$$+\frac{1}{2} N g^{\#}\left[\pi' \cdot \pi' - \frac{1}{2}(\operatorname{tr}\pi')^2\right] d\mu(g) - N\left[\mathrm{Ric}(g) - \frac{1}{2} R(g) g\right]^{\#} d\mu(g) +$$
$$+ (\mathrm{Hess}\, N + g\Delta N)^{\#} d\mu(g) + L_X \pi,$$

$$\text{Уравнения связи} \begin{cases} \mathcal{H}(g, \pi) = \left[\pi' \cdot \pi' - \frac{1}{2}(\operatorname{tr}\pi')^2 - R(g)\right] d\mu(g) = 0, \\ \mathcal{J}(g, \pi) = 2(\delta_g \pi) = -2\pi^{j}_{i|j} = 0. \end{cases}$$

Обратно, если i_λ *есть разбиение* $(V_4, {}^{(4)}g)$, *такое, что удовлетворяются записанные выше уравнения эволюции и связи, то* $^{(4)}g$ *удовлетворяет (вакуумным) уравнениям Эйнштейна.*

В формулировке этой теоремы используются следующие обозначения: $(\pi' \times \pi')^{ij} = (\pi')^{ik}(\pi')_k{}^j$; $\pi' \cdot \pi' = (\pi')^{ij}(\pi')_{ij}$; $\mathrm{Hess}\ N = N_{|i|j}$; $\Delta N = -g^{ij} N_{|i|j}$ и $L_X \pi = (L_X \pi') d\mu(g) + \pi'(\operatorname{div} X) d\mu(g)$ есть производная Ли тензорной плотности $\pi = \pi' d\mu(g)$; заметим, что $L_X d\mu(g) = (\operatorname{div} X) d\mu(g)$; Тензор Риччи для метрики $^{(4)}g$ обозначен как $\mathrm{Ric}(^{(4)}g)$, а для метрики g — как $\mathrm{Ric}\, g$; $R(g)$ — скалярная кривизна для метрики g. Тензор Эйнштейна для метрики g записывается в виде $\mathrm{Ein}(g) = \mathrm{Ric}(g) - \frac{1}{2}(g) g$.

Набросок доказательства теоремы 1 будет дан после теоремы 3.

Двенадцать уравнений эволюции первого порядка для (g, π) соответствуют шести уравнениям второго порядка $^{(4)}G_{ij}=0$, тогда как остальные четыре уравнения Эйнштейна $^{(4)}G^{00}=0$ и $^{(4)}G^0_i=0$ выступают в качестве уравнений связи. Точнее говоря, в координатах, задаваемых разбиением i_λ, $^{(4)}Z_\Sigma$ имеет компоненты $^{(4)}Z_c = (-N, 0)$. Если определить «поперечно-поперечную» и «поперечно-параллельную» проекции тензора Эйнштейна выражениями

$$^{(4)}G_{\perp\perp} = Z_\alpha Z_\beta\, {}^{(4)}G^{\alpha\beta} = N^2\, {}^{(4)}G^{00},$$
$$^{(4)}G^{\perp}{}_i = -Z_\alpha\, {}^{(4)}G^{\alpha}{}_i = N\, {}^{(4)}G^0{}_i,$$

то

$$\mathcal{H}(g, \pi) = -2\, {}^{(4)}G_{\perp\perp} d\mu(g), \qquad \mathcal{J}(g, \pi)_i = 2\, {}^{(4)}G_i^{\perp} d\mu(g).$$

Уравнения эволюции этой теоремы корректны, как показано в разд. 4.

В формулировке теоремы 1 длительность и сдвиг рассматриваются как независимо задаваемые величины. В формулировке «тонкого сэндвича» g и $\dot{g}$ рассматриваются как данные Коши, π выражается как функция $(\dot{g}, N, X)$, и затем из уравнений связи

$$\mathscr{H}(g, \pi(\dot{g}, N, X)) = 0,$$
$$\mathscr{J}(g, \pi(\dot{g}, N, X)) = 0$$

находятся N и X [144]. При линеаризации легко убедиться, что это не эллиптическая система, и поэтому, даже если ее и можно решить, здесь возникает ряд технических проблем, в частности неизбежная потеря регулярности. По этой причине большинство авторов отказалось от формулировки «тонкого сэндвича». Относительно других трудностей этой формулировки см. [55].

Важно обратить внимание на то, что многие члены в уравнениях эволюции АДМ комбинируются в виде производных Ли, и мы учли это в самой записи теоремы 1. Полезно также записать квадратичную алгебраическую часть $\partial\pi/\partial\lambda$ в виде

$$S_g(\pi, \pi) = -2\left\{\pi' \times \pi' - \frac{1}{2}(\operatorname{tr}\pi')\pi'\right\} d\mu(g) + \\ + \frac{1}{2} g^{\#} \left\{\pi' \cdot \pi' - \frac{1}{2}(\operatorname{tr}\pi')^2\right\} d\mu(g).$$

Это выражение представляет собой «пучок» метрики Де Витта, т. е. совокупность членов уравнения эволюции, квадратичных по π' (см. ниже, а также работу [83]). Таким образом, в уравнении эволюции для π отдельные члены можно интерпретировать следующим образом:

$$\frac{\partial\pi}{\partial\lambda} = NS_g(\pi, \pi) - N\,\mathrm{Ein}(g)^{\#}\, d\mu(g) + (\mathrm{Hess}\, N + g\Delta N)^{\#}\, d\mu(g) + L_X\pi.$$

геодезический «пучок» метрики Де Витта; вынуждающая «сила» потенциала скалярной кривизны; «противодействие» из-за непостоянного N; «сдвиговый» член

Дальнейшие сведения по геометрической интерпретации этого уравнения читатель найдет в работах [65, 83, 121—123].

Для интерпретации этих уравнений в терминах симплектической структуры кокасательного расслоения мы должны ввести следующие пространства. Обозначим через $\mathscr{M}$ пространство римановых C^∞-метрик на M и через $\mathscr{D} = \mathscr{D}(M)$ — группу диффеоморфизмов многообразия M. Через $\mathscr{M}^{s,p}$ при $s > n/p$ обозначим римановы метрики класса Соболева $W^{s,p}$; диффеоморфизмы и другие отображения, а также тензоры класса $W^{s,p}$ будут обозначаться аналогично. Однако для простоты обозначений мы ограничимся в этом разделе классом C^∞.

Пусть $T\mathscr{M} \approx \mathscr{M} \times S_2$ — касательное расслоение пространства $\mathscr{M}$, где, как и раньше, S_2— пространство 2-ковариантных симметричных тензорных полей класса C^∞ и S^2_d —

пространство 2-контравариантных тензорных плотностей класса C^∞ на M. Введем $T^*\mathcal{M} \approx \mathcal{M} \times S_d^2 = \{(g, \pi) | g \in \mathcal{M}, \pi \in S_d^2\}$. Будем рассматривать $T^*\mathcal{M}$ как «L_2-кокасательное расслоение к $\mathcal{M}$». Для $k \in T_g\mathcal{M} \approx S_2$, $\pi \in T_g^*\mathcal{M} \approx S_d^2$ имеется, как объяснялось выше, естественное L_2-спаривание

$$(\pi, k)_{L_2} = \int_M \pi \cdot k.$$

При таком определении $T^*\mathcal{M}$ является подрасслоением «истинного» кокасательного расслоения. Поскольку $T^*\mathcal{M}$ есть открытое множество в $S_2 \times S_d^2$, касательным пространством к $T^*\mathcal{M}$ в точке $(g, \pi) \in T^*\mathcal{M}$ является $T_{(g,\pi)}(T^*\mathcal{M}) \approx S_2 \times S_d^2$.

Мы покажем теперь, что $T^*\mathcal{M}$ обладает естественной симплектической структурой, в которой уравнения эволюции теоремы 1 являются гамильтоновыми. Для включения в эту схему функции длительности и векторного поля сдвига необходимо ввести понятие обобщенной гамильтоновой системы.

Определим на $T^*\mathcal{M}$ глобально постоянную симплектическую структуру

$$\Omega = \Omega_{(g,\pi)} : T_{(g,\pi)}(T^*\mathcal{M}) \times T_{(g,\pi)}(T^*\mathcal{M}) \to \mathbb{R}$$

следующим образом: для $(h_1, \omega_1), (h_2, \omega_2) \in T_{(g,\pi)}(T^*\mathcal{M}) = S_2 \times S_d^2$,

$$\Omega_{(g,\pi)}((h_1, \omega_1), (h_2, \omega_2)) = \int_M \omega_2 h_1 - \omega_1 \cdot h_2.$$

Пусть $J = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix} : S_d^2 \times S_2 \to S_2 \times S_d^2$ определяется следующим образом: $(\omega, h) \mapsto J\begin{pmatrix} \omega \\ h \end{pmatrix} = \begin{pmatrix} h \\ -\omega \end{pmatrix}$, так что $J^{-1} = \begin{pmatrix} 0 & -I \\ I & 0 \end{pmatrix} : S_2 \times S_d^2 \to S_d^2 \times S_2$, $(h, \omega) \mapsto (-\omega, h)$.

Тогда

$$\Omega((h_1, \omega_1), (h_2, \omega_2)) = \int_M \left\langle J^{-1}\begin{pmatrix} h_1 \\ \omega_1 \end{pmatrix}, (h_2, \omega_2) \right\rangle.$$

Вскоре мы вернемся к матрице J.

Пусть $C^\infty = C^\infty(M; \mathbb{R})$ означает гладкие функции на M с действительными значениями; C_d^∞ — гладкие скалярные плотности на M; $\mathcal{X}$ — гладкие векторные поля на M; Λ_d^1 — гладкие плотности линейных форм на M. Рассмотрим функции

$$\mathcal{H} : T^*\mathcal{M} \mapsto C_d^\infty;\ (g, \pi) \mapsto \mathcal{H}(g, \pi) = \\ = \left[\pi' \cdot \pi' - \frac{1}{2}(\operatorname{tr}\pi')^2 - R(g)\right] d\mu(g);$$

$$\mathcal{J} = 2\delta : T^*\mathcal{M} \mapsto \Lambda_d^1; \quad (g, \pi) \mapsto 2(\delta_g \pi) = -2\pi_i{}^j{}_{|j};$$

$$\Phi = (\mathcal{H}, \mathcal{J}) : T^*\mathcal{M} \mapsto C_d^\infty \times \Lambda_d^1; \quad (g, \pi) \mapsto (\mathcal{H}(g, \pi), \mathcal{J}(g, \pi)).$$

На этом этапе необходимо вычислить производные $\mathcal{H}$, $\mathcal{J}$ и Φ и сопряженные им естественным образом величины. Результаты вычислений представляются в виде следующего предложения.

Предложение 2

Если положить $(g, \pi) \in T^*\mathcal{M}$, $(h, \omega) \in T_{(g,\pi)}(T^*\mathcal{M}) = S_2 \times S_d^2$ *и* $(N, X) \in C^\infty \times \mathcal{X}$, *то производные функций* $\mathcal{H}$, $\mathcal{J}$, Φ

$$\mathrm{D}\mathcal{H}(g, \pi): S_2 \times S_d^2 \to C_d^\infty,$$
$$\mathrm{D}\mathcal{J}(g, \pi): S_2 \times S_d^2 \to \Lambda_d^1,$$
$$\mathrm{D}\Phi(g, \pi): S_2 \times S_d^2 \to C_d^\infty \times \Lambda_d^1$$

и естественным образом сопряженные им величины

$$[\mathrm{D}\mathcal{H}(g, \pi)]^*: C^\infty \to S_d^2 \times S_2,$$
$$[\mathrm{D}\mathcal{J}(g, \pi)]^*: \mathcal{X} \to S_d^2 \times S_2,$$
$$[\mathrm{D}\Phi(g, \pi)]^*: C^\infty \times \mathcal{X} \to S_d^2 \times S_2$$

даются следующими выражениями:

$$\mathrm{D}\mathcal{H}(g, \pi)\cdot(h, \omega) = -S_g(\pi, \pi)\cdot h + \\ + [\mathrm{Ein}(g)\cdot h - (\delta\delta h + \Delta \operatorname{tr} h)]\, d\mu(g) + 2\left[(\pi') - \frac{1}{2}(\operatorname{tr}\pi')\, g\right]\cdot\omega;$$

$$[\mathrm{D}\mathcal{H}(g, \pi)]^*\cdot N = \\ = \left\{[-NS_g(\pi, \pi) + (N\,\mathrm{Ein}(g) - (\mathrm{Hess}\, N + g\Delta N))^\#] \otimes d\mu(g), \right. \\ \left. 2N\left[(\pi')^\flat - \frac{1}{2}(\operatorname{tr}\pi')\, g\right]\right\};$$

$$\mathrm{D}\mathcal{J}(g, \pi)\cdot(h, \omega) = -2\left[\omega_i{}^j{}_{|j} + h_{ik}\pi^{kj}{}_{|j} + \pi^{jl}\left(h_{ij|l} - \frac{1}{2} h_{jl|i}\right)\right];$$

$$[\mathrm{D}\mathcal{J}(g, \pi)]^*\cdot X = (-L_X\pi,\ L_X g);$$

$$\mathrm{D}\Phi(g, \pi)\cdot(h, \omega) = (\mathrm{D}\mathcal{H}(g, \pi)\cdot(h, \omega),\ \mathrm{D}\mathcal{J}(g, \pi)\cdot(h, \omega));$$

$$[\mathrm{D}\Phi(g, \pi)]^*\cdot(N, X) = [\mathrm{D}\mathcal{H}(g, \pi)]^*\cdot N + [\mathrm{D}\mathcal{J}(g, \pi)]^*\cdot X = \\ = \left\{[-NS_g(\pi, \pi) + (N\,\mathrm{Ein}(g) - (\mathrm{Hess}\, N + g\Delta N))^\#] \otimes d\mu(g) - \right. \\ \left. -L_X\pi,\ 2N\left[(\pi')^\flat - \frac{1}{2}(\operatorname{tr}\pi')\, g + L_X g\right]\right\}.$$

Доказательство состоит в довольно длинном, но прямом вычислении.

Как показано в работе [7], уравнения эволюции теоремы 1 являются гамильтоновыми уравнениями с гамильтонианом $N\mathcal{H} + X\cdot\mathcal{J}$, т. е.

$$\frac{\partial g}{\partial\lambda} = \frac{\delta}{\delta\pi}(N\mathcal{H} + X\cdot\mathcal{J}),$$
$$\frac{\partial\pi}{\partial\lambda} = -\frac{\delta}{\delta g}(N\mathcal{H} + X\cdot\mathcal{J}).$$

Используя симплектическую структуру на $T^*\mathcal{M}$, задаваемую ма-

трицей

$$J=\begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}: S_d^2\times S_2 \to S_2\times S_d^2,\ (\omega,\ h)\mapsto J\begin{pmatrix}\omega\\ h\end{pmatrix}=\begin{pmatrix} h\\ -\omega\end{pmatrix},$$

и соответствие

$$\left(\frac{\delta}{\delta g}(N\mathscr{H}+X\cdot\mathscr{Y}),\ \frac{\delta}{\delta\pi}(N\mathscr{H}+X\cdot\mathscr{Y})\right)=[\mathrm{D}\Phi(g,\ \pi)]^*\cdot\begin{pmatrix}N\\ X\end{pmatrix},$$

гамильтоновы уравнения теоремы 1 можно представить в весьма компактной форме.

Теорема 3

Система уравнений Эйнштейна, задаваемая уравнениями эволюции и уравнениями связи в теореме 1, может быть представлена в виде

$$\text{Уравнения эволюции } \frac{\partial}{\partial\lambda}\begin{pmatrix}g\\ \pi\end{pmatrix}=J\circ[\mathrm{D}\Phi(g,\ \pi)]^*\cdot\begin{pmatrix}N\\ X\end{pmatrix}$$

$$\text{Уравнения связи } \Phi(g,\ \pi)=(\mathscr{H}(g,\ \pi),\ \mathscr{Y}(g,\ \pi))=0,$$

где $\{N,\ X\}$ суть функция длительности и векторное поле сдвига, ассоциированные с данным разбиением, а

$$[\mathrm{D}\Phi(g,\ \pi)]^*\cdot\begin{pmatrix}N\\ X\end{pmatrix}$$

определяется в предложении 2.

Набросок доказательства теорем 1 и 3

Лагранжева плотность, порождающая уравнение Эйнштейна для пустого пространства имеет вид

$$\mathscr{L}_{\text{грав}}({}^{(4)}g)=\frac{1}{16\pi}R({}^{(4)}g)\,d\mu({}^{(4)}g),$$

где $d\mu({}^{(4)}g)=(-\det{}^{(4)}g)^{1/2}d^4x=N(\det g)^{1/2}d^3x d\lambda=N d\lambda d\mu(g)$. Из вычислительной части доказательства, которую мы здесь опускаем, видно, что $\mathscr{L}_{\text{грав}}$ можно записать в следующей (3+1)-мерной форме (см. уравнения 7-3.13 в [7] и уравнения 21-90 в [144]):

$$16\pi\mathscr{L}_{\text{грав}}({}^{(4)}g)=NR({}^{(4)}g)\,d\mu(g)\,d\lambda=$$

$$=\left[\pi^{ij}\frac{\partial g_{ij}}{\partial\lambda}-N\mathscr{H}(g,\ \pi)-X\cdot\mathscr{Y}(g,\ \pi)\right]d\lambda-$$

$$-2\left[\pi^i_j X^j-\frac{1}{2}X^i\operatorname{tr}\pi+(\operatorname{grad}N)^i\,d\mu(g)\right]_{,i}d\lambda-\left(\frac{\partial}{\partial\lambda}\operatorname{tr}\pi\right)d\lambda.$$

Здесь i_λ есть разбиение многообразия V_4, так что V_4 можно отождествить с $I\times M$. Заметим, что наше определение $\pi=\pi' d\mu(g)=$

$=\pi'(\det g)^{1/2}d^3x=\pi^{\text{АДМ}}\,d^3x$ содержит множитель d^3x, который дополняет $(\det g)^{1/2}$ до элемента объема на M. Подобным же образом элемент объема $d\mu(^{(4)}g)$ включает $d^4x=d^3x\,d\lambda$, чем и объясняется наличие общего множителя $d\lambda$.

Построим на M векторную плотность $\beta=\beta^i=2[\pi^i{}_j X^j-\frac{1}{2}X^i\operatorname{tr}\pi+ +(\operatorname{grad} N)^i\,d\mu(g)]$ и заметим, что $\beta^i{}_{;i}=\beta^i{}_{|i}=\operatorname{div}\beta$. Действие для гравитации можно записать в виде

$$16\pi S_{\text{грав}}(^{(4)}g)=16\pi\int_{V_4}\mathscr{L}_{\text{грав}}(^{(4)}g)=$$
$$=16\pi\int_I d\lambda\int_M\left[\pi\cdot\frac{\partial g}{\partial\lambda}-N\mathscr{H}(g,\pi)-X\cdot\mathscr{J}(g,\pi)\right]+$$
$$+16\pi\int_I d\lambda\int_M\left(\operatorname{div}\beta-\frac{\partial}{\partial\lambda}\operatorname{tr}\pi\right).$$

Поскольку интегрирование по M обращает член с $\operatorname{div}\beta$ в нуль, а член с полной производной по времени

$$\int_{I=[a,\,b]}d\lambda\int_M\frac{\partial}{\partial\lambda}\operatorname{tr}\pi=\int_M(\operatorname{tr}\pi)_{\lambda=b}-\int_M(\operatorname{tr}\pi)_{\lambda=a}$$

можно опустить как постоянную, которая не войдет в вариацию $S_{\text{грав}}$, имеем

$$16\pi S_{\text{грав}}(^{(4)}g)=16\pi\int_I d\lambda\int_M\left(\pi\cdot\frac{\partial g}{\partial\lambda}-N\mathscr{H}-X\cdot\mathscr{J}\right)=$$
$$=16\pi\int_I d\lambda\int_M\left[\pi\cdot\frac{\partial g}{\partial\lambda}-\Phi(g,\pi)\cdot\begin{pmatrix}N\\X\end{pmatrix}\right].$$

Вариация действия по $^{(4)}g$ в направлении $^{(4)}h$, исчезающая на $\{a\}\times M$ и $\{b\}\times M$, индуцирует вариацию (h,ω) пары (g,π), которая также исчезает на каждом концевом многообразии $\{a\}\times M$ и $\{b\}\times M$. Таким образом, из условия экстремума действия при произвольной вариации (h,ω), исчезающей на концевых многообразиях $\{a\}\times M$ и $\{b\}\times M$, получим

$$0=16\pi\,dS_{\text{грав}}(^{(4)}g)\cdot{}^{(4)}h=16\pi\int_I d\lambda\int_M\left(\omega\cdot\frac{\partial g}{\partial\lambda}+\pi\cdot\frac{\partial h}{\partial\lambda}\right)-$$
$$-16\pi\int_I d\lambda\int_M\left\langle D\Phi(g,\pi)\cdot(h,\omega),\begin{pmatrix}N\\X\end{pmatrix}\right\rangle=$$
$$=16\pi\int_I d\lambda\int_M\left(\omega\cdot\frac{\partial g}{\partial\lambda}-\frac{\partial\pi}{\partial\lambda}\cdot h\right)+16\pi\left[\int_M(\pi\cdot h)_{\lambda=b}-\int_M(\pi\cdot h)_{\lambda=a}\right]-$$
$$-16\pi\int_I d\lambda\int_M\left\langle(h,\omega),[D\Phi(g,\pi)]^*\cdot\begin{pmatrix}N\\X\end{pmatrix}\right\rangle=$$
$$=16\pi\int_I d\lambda\int_M\left\langle(h,\omega),\left[\left(-\frac{\partial\pi}{\partial\lambda},\frac{\partial g}{\partial\lambda}\right)-[D\Phi(g,\pi)]^*\cdot\begin{pmatrix}N\\X\end{pmatrix}\right]\right\rangle,$$

где член, содержащий полную производную по времени

$$\int_I d\lambda \int_M \frac{\partial(\pi\cdot h)}{\partial\lambda},$$

обращается в нуль при интегрировании по переменной λ ввиду исчезновения h на концевых многообразиях. Из произвольности вариации (h, ω) следует, что

$$\left(-\frac{\partial\pi}{\partial\lambda}, \frac{\partial g}{\partial\lambda}\right) = [D\Phi(g, \pi)]^* \cdot \binom{N}{X},$$

а потому

$$J\begin{pmatrix} -\frac{\partial\pi}{\partial\lambda} \\ \frac{\partial g}{\partial\lambda} \end{pmatrix} = \begin{pmatrix} \frac{\partial g}{\partial\lambda} \\ \frac{\partial\pi}{\partial\lambda} \end{pmatrix} = J \circ [D\Phi(g, \pi)]^* \cdot \binom{N}{X}. \blacksquare$$

Теперь мы обсудим некоторые дополнительные детали гамильтоновой структуры сопряженного уравнения в теореме 3.

Пусть $F: T^*\mathcal{M} \to \mathbb{R}$ есть функция на $T^*\mathcal{M}$ с действительными значениями, задаваемая плотностью $\mathcal{F}: T^*\mathcal{M} \to C_d^\infty$, т. е.

$$F(g, \pi) = \int_M \mathcal{F}(g, \pi).$$

Тогда *гамильтоново векторное поле* функции F

$$X_F: T^*\mathcal{M} \to T(T^*\mathcal{M})$$

определяется равенством

$$dF(g, \pi)\cdot(h, \omega) = \Omega(X_F(g, \pi), (h, \omega)),$$

где Ω — симплектическая структура на $T^*\mathcal{M}$.

Предложение 4

Гамильтоново векторное поле X_F дается выражением

$$X_F(g, \pi) = J \circ [D\mathcal{F}(g, \pi)]^* \cdot 1.$$

Доказательство

$$\Omega(X_F(g, \pi), (h, \omega)) = -\int \langle X_F(g, \pi), J^{-1}(h, \omega)\rangle,$$

и, таким образом,

$$dF(g, \pi)\cdot(h, \omega) = \int D\mathcal{F}(g, \pi)\cdot(h, \omega) = \int \langle [D\mathcal{F}(g, \pi)]^* \cdot 1, (h, \omega)\rangle =$$

$$= -\int \langle J \circ [D\mathcal{F}(g, \pi)]^* \cdot 1, J^{-1}(h, \omega)\rangle$$

$(J^* = -J)$

$$= \Omega\{J \circ [D\mathcal{F}(g, \pi)]^* \cdot 1, (h, \omega)\}. \blacksquare$$

В частности, если $F=\int N\mathscr{H}+X\cdot\mathscr{Y}=\int\langle(N, X), \Phi\rangle$, то

$$X_F(g, \pi)=J\circ[\mathrm{D}(N\mathscr{H}+X\cdot\mathscr{Y})]^*\cdot 1=J\circ[\mathrm{D}\Phi(g, \pi)]^*\cdot\begin{pmatrix}N\\X\end{pmatrix}.$$

Отсюда видно, что эйнштейновские уравнения эволюции являются гамильтоновыми уравнениями на симплектическом многообразии $T^*\mathscr{M}$ с плотностью гамильтониана $N\mathscr{H}+X\cdot\mathscr{Y}$.

Допустим теперь, что $F_1, F_2: T^*\mathscr{M}\to\mathbb{R}$ являются функциями на $T^*\mathscr{M}$ с действительными значениями, задаваемыми плотностями $\mathscr{F}_1$ и $\mathscr{F}_2$ соответственно. Тогда их скобка Пуассона

$$\{F_1, F_2\}: T^*\mathscr{M}\to\mathbb{R}$$

определяется как

$$\{F_1, F_2\}(g, \pi)=\Omega(X_{F_1}(g, \pi), X_{F_2}(g, \pi)),$$

где X_F — гамильтоново векторное поле для F.

Предложение 5

Для скобки Пуассона $\{F_1, F_2\}$, определенной выше, справедливо выражение

$$\{F_1, F_2\}(g, \pi)=\int\langle[\mathrm{D}_g\mathscr{F}_1(g, \pi)]^*\cdot 1, [\mathrm{D}_\pi\mathscr{F}_2(g, \pi)]^*\cdot 1\rangle-$$
$$-\int\langle[\mathrm{D}_\pi\mathscr{F}_1(g, \pi)]^*\cdot 1, \mathrm{D}_g\mathscr{F}_2(g, \pi)\cdot 1\rangle.$$

Доказательство

$$\{F_1, F_2\}(g, \pi)=\Omega(X_{F_1}(g, \pi), X_{F_2}(g, \pi))=$$
$$=-\int\langle X_{F_1}(g, \pi), J^{-1}\circ X_{F_2}(g, \pi)\rangle=$$
$$=-\int\langle J\circ[\mathrm{D}\mathscr{F}_1(g, \pi)]^*\cdot 1, J^{-1}\circ J\circ[\mathrm{D}\mathscr{F}_2(g, \pi)]^*\cdot 1\rangle=$$
$$=-\int\langle[\mathrm{D}\mathscr{F}_1(g, \pi)]^*\cdot 1, J^*\circ[\mathrm{D}\mathscr{F}_2(g, \pi)]^*\cdot 1\rangle=$$
$$=\int\langle[\mathrm{D}\mathscr{F}_1(g, \pi)]^*\cdot 1, J\circ[\mathrm{D}\mathscr{F}_2(g, \pi)]^*\cdot 1\rangle=$$
$$=\int\langle([\mathrm{D}_g\mathscr{F}_1(g, \pi)]^*\cdot 1, [\mathrm{D}_\pi\mathscr{F}_1(g, \pi)]^*\cdot 1)\times$$
$$\times\begin{pmatrix}[\mathrm{D}_\pi\mathscr{F}_2(g, \pi)]^*\cdot 1\\-[\mathrm{D}_g\mathscr{F}_2(g, \pi)]^*\cdot 1\end{pmatrix}\rangle=$$
$$=\int\langle[\mathrm{D}_g, \mathscr{F}_1(g, \pi)]^*\cdot 1, [\mathrm{D}_\pi\mathscr{F}_2(g, \pi)]^*\cdot 1\rangle-$$
$$-\int\langle[\mathrm{D}_\pi\mathscr{F}_1(g, \pi)]^*\cdot 1, [\mathrm{D}_g\mathscr{F}_2(g, \pi)]^*\cdot 1\rangle.$$

Этот результат в «физических обозначениях» можно записать в виде

$$\{F_1, F_2\}=\int\left(\frac{\delta\mathscr{F}_1}{\delta g}\frac{\delta\mathscr{F}_2}{\delta\pi}-\frac{\delta\mathscr{F}_1}{\delta\pi}\frac{\delta\mathscr{F}_2}{\delta g}\right).$$

Теперь рассмотрим случай, когда

$$F_2 = \int N\mathscr{H} + X\cdot\mathscr{Y}.$$

Тогда из приведенного выше доказательства вытекает, что

$$\{F,\ N\mathscr{H}+X\cdot\mathscr{Y}\}(g,\pi) = \int \langle [D\mathscr{F}(g,\pi)]^*\cdot 1,\ J\circ[D(N\mathscr{H}+X\cdot\mathscr{Y})]^*\cdot 1\rangle$$

$$= \int \left\langle [D\mathscr{F}(g,\pi)]^*\cdot 1,\ J\circ[D\Phi(g,\pi)]^*\cdot\begin{pmatrix}N\\X\end{pmatrix}\right\rangle =$$

$$= \int D\mathscr{F}(g,\pi)\cdot\begin{pmatrix}\frac{\partial g}{\partial\lambda}\\ \frac{\partial\pi}{\partial\lambda}\end{pmatrix} = \int \frac{d}{d\lambda}\mathscr{F}(g,\pi) =$$

$$= \frac{d}{d\lambda}F(g,\pi).$$

Это означает следующее. Пусть $(g(\lambda),\ \pi(\lambda))$ — решение эволюционных уравнений Эйнштейна с длительностью и сдвигом $N(\lambda)$, $X(\lambda)$. Пусть $F(\lambda)=F(g(\lambda),\ \pi(\lambda))$. Тогда

$$\frac{dF}{d\lambda} = \{F,\ N\mathscr{H}+X\cdot\mathscr{Y}\}.$$

Таким образом, скобка Пуассона с гамильтонианом $\int N\mathscr{H}+ +X\cdot\mathscr{Y}$ порождает, как и ожидалось, производные $F(g(\lambda),\ \pi(\lambda))$ по λ, где $(g(\lambda),\ \pi(\lambda))$ есть поток с начальными данными $(g(0),\ \pi(0))$ и с длительностью и сдвигом $(N(\lambda),\ X(\lambda))$.

Запись уравнений Эйнштейна, приведенную в теореме 3, можно на самом деле обобщить так, чтобы она могла включать теории полей, взаимодействующих с гравитацией. Эта обобщенная форма лежит в основе ковариантной формулировки гамильтоновых систем [90, 92, 53, 121—123]. Например, каноническую формулировку ковариантного скалярного волнового уравнения $\Box\phi=m^2\phi+F'(\phi)$ в пространстве-времени $(V_4=I\times M,\ {}^{(4)}g)$ можно получить через общие длительность и сдвиг следующим образом.

Рассмотрим гамильтониан для скалярного поля

$$\mathscr{H}(g;\ \phi,\pi) = \left\{\frac{1}{2}\left[(\pi'_\phi)^2 + |\nabla\phi|^2 + m^2\phi^2\right] + F(\phi)\right\}d\mu(g)$$

(в этом примере фоновая метрика рассматривается как заданная, но без уточнения ее конкретного вида). Мы можем построить дважды ковариантную симметричную тензорную плотность $\mathscr{T}$, варьируя $\mathscr{H}(g;\ \phi,\ \pi_\phi)$ по g:

$$\mathscr{T} = -2[D_g\mathscr{H}(g;\ \phi,\ \pi_\phi)]^*\cdot 1,$$

и плотность линейной формы $\mathscr{Y}(\phi,\ \pi_\phi)$, пользуясь соотношением

$$\int\langle X,\ \mathscr{Y}(\phi,\ \pi_\phi)\rangle = \int\langle\pi_\phi,\ L_X\phi\rangle,$$

откуда $\mathscr{Y}(\phi, \pi_\phi)=\pi_\phi \cdot d\phi$. Это равенство представляет $\mathscr{Y}$ как сохраняющуюся величину для координатной группы инвариантности на M [83, 84]. Если мы положим $\Phi=(\mathscr{H}, \mathscr{Y})$, то гамильтоновы уравнения для ϕ в общем разбиении пространства с длительностью N и сдвигом X примут вид

$$\frac{\partial}{\partial\lambda}\begin{pmatrix}\phi \\ \pi_\phi\end{pmatrix} = J\circ[\mathrm{D}\Phi(g;\ \phi,\ \pi_\phi)]^{*}\cdot\begin{pmatrix}N \\ X\end{pmatrix},$$

т. е. точно такой же, как для общей теории относительности. Вычисления показывают, что эта система эквивалентна приведенному выше скалярному волновому уравнению. Здесь $\mathrm{D}\Phi(g;\ \phi,\ \pi_\phi)$ есть производная Φ по скалярному полю и его каноническому импульсу π_ϕ.

Если мы вводим взаимодействие скалярного поля с гравитацией, рассматривая скалярное поле как источник, то уравнение для гравитационного импульса $\partial\pi/\partial\lambda$ в теоремах 1 и 3 изменяется на дополнительный член $(^1/_2)N\mathscr{T}$, а уравнение для $\partial g/\partial\lambda$ остается неизменным. Уравнения связи принимают вид

$$\mathscr{H}_{\text{грав}}(g,\ \pi)+\mathscr{H}_{\text{скал}}(g;\ \ \phi,\ \pi_\phi)=0, \qquad \mathscr{Y}_{\text{грав}}(g,\ \pi)+\mathscr{Y}_{\text{скал}}(\phi,\ \pi\ )=0.$$

И вообще, если рассматриваются полный гамильтониан $\mathscr{H}_T=\mathscr{H}_{\text{грав}}+\mathscr{H}_{\text{полей}}$ и тензор полного потока $\mathscr{Y}_T=\mathscr{Y}_{\text{грав}}+\mathscr{Y}_{\text{полей}}$ и если взаимодействие негравитационных полей с гравитационным полем является связью без производных (т. е. взаимодействие минимально), то уравнения с этим взаимодействием имеют вид

$$\frac{\partial}{\partial\lambda}\begin{pmatrix} g \\ \pi \\ \phi_A \\ \pi^A\end{pmatrix} = J\circ[\mathrm{D}\Phi_{\mathrm{T}}(g,\ \pi;\ \phi_A,\ \pi^A)]^{*}\cdot\begin{pmatrix}N \\ X \\ \psi\end{pmatrix};$$

$$\Phi(g,\ \pi;\ \ \phi_A,\ \pi^A)=(\Phi_{\mathrm{T}}(g,\ \ \pi;\ \phi_A,\ \pi^A),\ \Phi_{\text{вырожд}}(g,\ \pi;\ \ \phi_A,\ \pi^A))=0.$$

Здесь под ϕ_A имеются в виду все негравитационные динамические поля, под π^A— их сопряженные импульсы, под $\Phi_{\text{вырожд}}=0$ — дополнительные связи, возникающие вследствие вырождения [4] $\mathscr{L}$, а ψ суть соответствующие нединамические (вырожденные) поля. Эти результаты обеспечивают единую ковариантную гамильтонову формулировку общей теории относительности в сочетании с другими лагранжевыми полевыми теориями и фактически позволяют формально перейти от случая пустого пространства к случаю связи без производных. Приведенное доказательство того, что описание полей, взаимодействующих с гравитацией, может быть дано в "L_2-сопряженном формализме», основано на результатах Кухаржа. В серии работ [119—125], составивших веху в этой области (первые шаги в которой были сделаны Дираком, см. [79] и ссылки, приведенные там), Кухарж дал разработанную в деталях каноническую формулировку ковариантных полевых теорий. Реализацию этой

формулировки для полей Янга — Миллса можно найти в работах Армса [2, 3].

Формализм, изложенный здесь, можно распространить на случай некомпактного многообразия M. В этом случае возникает много технических проблем, но принципиальное различие только одно: скорость спадания асимптотически плоской метрики недостаточна для того, чтобы было возможно интегрирование по частям. Это привело Редже и Тейтельбойма [161] к выводу, что настоящий гамильтониан, действительно порождающий эволюционные уравнения, содержит дополнительный член с поверхностным интегралом, соответствующий массе. Таким образом, в асимптотически-плоском случае после наложения связи $\Phi=0$ эту массу можно истолковать как «истинный» генератор уравнений эволюции. Эти идеи обсуждаются в работе [54].

2. МНОГООБРАЗИЕ СВЯЗЕЙ

Пусть через $\mathscr{C}_{\mathscr{H}}=\{(g, \pi)\in T^*\mathscr{M}|\mathscr{H}(g, \pi)=0\}$ обозначено множество решений гамильтоновой связи, и пусть

$$\mathscr{C}_\delta=\{(g, \pi)\in T^*\mathscr{M}\mid \mathscr{J}(g, \pi)=-2\pi_i{}^j{}_{|j}=0\}$$

означает множество решений дивергенциальной связи. Таким образом, $\mathscr{C}=\mathscr{C}_{\mathscr{H}}\cap\mathscr{C}_\delta\subset T^*\mathscr{M}$ есть множество связей для системы Эйнштейна в пустом пространстве. Отметим два следующих важных факта относительно $\mathscr{C}_{\mathscr{H}}\cap\mathscr{C}_\delta$:

1) Уравнения эволюции сохраняют эти связи при любом выборе функции течения времени и векторного поля сдвига.

2) $\mathscr{C}_{\mathscr{H}}\cap\mathscr{C}_\delta$ является в общем случае гладким подмногообразием в $T^*\mathscr{M}$.

С пространственно-временной точки зрения сохранение связей эквивалентно свернутым тождествам Бианки, которые являются дифференциальными тождествами, порожденными ковариантностью четырехмерных уравнений поля. Это сохранение связей во времени необходимо для согласованности уравнений эволюции и связей.

Структура многообразия на $\mathscr{C}_{\mathscr{H}}\cap\mathscr{C}_\delta$, интересная сама по себе, как мы увидим далее, является также ключом к пониманию устойчивости линеаризации полевых уравнений.

Для начала заметим, что гамильтониан и функции импульса ковариантны по отношению к $\mathscr{D}(M)$, бесконечномерной калибровочной группе диффеоморфизмов многообразия M. Иначе говоря, для любого $\eta\in\mathscr{D}(M)$ и $(g, \pi)\in T^*\mathscr{M}$

$$\mathscr{H}(\eta^*g, \eta^*\pi)=\eta^*\mathscr{H}(g, \pi),$$
$$\mathscr{J}(\eta^*g, \eta^*\pi)=\eta^*\mathscr{J}(g, \pi)$$

и, следовательно, $\Phi(\eta^*g, \eta^*\pi)=\eta^*\Phi(g, \pi)$.

Здесь η^* — обычное увлечение тензоров.

Если η_λ— некоторая кривая на $\mathscr{D}(M)$, причем η_0 есть единица группы, то, вводя векторное поле X как

$$X = \left(\frac{d}{d\lambda}\eta_\lambda\right)_{\lambda=0},$$

дифференцируя приведенные выше соотношения ковариантности по λ и положив $\lambda=0$, получим инфинитезимальную версию этих соотношений:

$$D\mathscr{H}(g, \pi)\cdot(L_X g, L_X\pi) = L_X[\mathscr{H}(g, \pi)],$$
$$D\mathscr{Y}(g, \pi)\cdot(L_X g, L_X\pi) = L_X[\mathscr{Y}(g, \pi)]$$

и, следовательно,

$$D\Phi(g, \pi)\cdot(L_X g, L_X\pi) = L_X[\Phi(g, \pi)].$$

Подобные же равенства возникают из калибровочной инвариантности полей Янга — Миллса.

Скорость изменения $\mathscr{H}$ и $\mathscr{Y}$ вдоль решения уравнения эволюции при функциях длительности и сдвига общего вида определяется следующей теоремой. Производная Ли в приведенных в ней уравнениях обусловлена уже обсуждавшейся инфинитезимальной ковариантностью.

Теорема 6

Пусть $(g(\lambda), \pi(\lambda))$ — некоторое решение эволюционных уравнений Эйнштейна

$$\frac{\partial}{\partial\lambda}\begin{pmatrix} g \\ \pi \end{pmatrix} = J\circ[D\Phi(g, \pi)]^*\cdot\begin{pmatrix} N \\ X \end{pmatrix}$$

при произвольной длительности $N(\lambda)$ и произвольном сдвиге $X(\lambda)$. Тогда $(\mathscr{H}(\lambda), \mathscr{Y}(\lambda)) = \{\mathscr{H}(g(\lambda), \pi(\lambda)), \mathscr{Y}(g(\lambda), \pi(\lambda))\}$ удовлетворяют следующей системе уравнений:

$$\frac{d\mathscr{H}}{d\lambda} = \frac{1}{N}\operatorname{div}(N^2\mathscr{Y}) + L_X\mathscr{H},$$
$$\frac{d\mathscr{Y}}{d\lambda} = (dN)\mathscr{H} + L_X\mathscr{Y}.$$

Если для некоторого λ_0 в области существования этого решения $(g(\lambda_0), \pi(\lambda_0)) = (g_0, \pi_0)\in\mathscr{C}_{\mathscr{H}}\cap\mathscr{C}_\delta$ (т. е. $\Phi(g_0, \pi_0)=0$), то $(g(\lambda), \pi(\lambda))\in\mathscr{C}_{\mathscr{H}}\cap\mathscr{C}_\delta$ при всех λ, для которых решение существует.

Замечание. Из теорем единственности (см. разд. 3) вытекает, что решение уравнений эволюции должно полностью лежать в $\mathscr{C}_{\mathscr{H}}\cap\mathscr{C}_\delta$, если оно пересекает $\mathscr{C}_{\mathscr{H}}\cap\mathscr{C}_\delta$.

Доказательство. Используем инфинитезимальную ковариант-

ность $\mathscr{H}$ следующим образом:

$$\frac{d\mathscr{H}(g,\pi)}{d\lambda} = \mathrm{D}\mathscr{H}(g,\ \pi)\cdot\left(\frac{\partial g}{\partial\lambda},\frac{\partial\pi}{\partial\lambda}\right) =$$
$$= \mathrm{D}\mathscr{H}(g,\pi)\cdot\left\{J\circ[\mathrm{D}\Phi(g,\ \pi)]^*\cdot\begin{pmatrix}N\\X\end{pmatrix}\right\} =$$
$$= \mathrm{D}\mathscr{H}(g,\ \pi)\cdot\{J\circ[(\mathrm{D}\mathscr{H}(g,\ \pi))^*\cdot N + (\mathrm{D}\mathscr{Y}(g,\ \pi))^*\cdot X]\} =$$
$$= \mathrm{D}\mathscr{H}(g,\ \pi)\cdot\{J\circ[(\mathrm{D}\mathscr{H}(g,\ \pi))^*\cdot N + (-L_X\pi,\ L_X g)]\} =$$
$$= \mathrm{D}_g\mathscr{H}(g,\ \pi)\{[\mathrm{D}_\pi\mathscr{H}(g,\ \pi)]^*\cdot N\} -$$
$$- \mathrm{D}_\pi\mathscr{H}(g,\ \pi)\cdot\{[\mathrm{D}_g\mathscr{H}(g,\ \pi)]^*\cdot N\} + L_X\mathscr{H}(g,\ \pi).$$

Первые два члена в выражении для $\partial\mathscr{H}/\partial\lambda$ требуют довольно трудоемких вычислений. Результат имеет вид

$$\mathrm{D}_g\mathscr{H}\cdot[(\mathrm{D}_\pi\mathscr{H})^*\cdot N] - \mathrm{D}_\pi\mathscr{H}\cdot[(\mathrm{D}_g\mathscr{H})^*\cdot N] = -\frac{1}{N}\delta[N^2\delta(2\pi)].$$

Отсюда мы приходим к следующему выражению:

$$\frac{d\mathscr{H}}{d\lambda} = -\frac{1}{N}\delta[N^2\delta(2\pi)] + L_X\mathscr{H} = \frac{1}{N}\operatorname{div}(N^2\mathscr{Y}) + L_X\mathscr{H}.$$

Эволюционное уравнение для $\mathscr{Y}(g,\pi)$ следует из инфинитезимальной ковариантности $\Phi(g,\pi)$ следующим образом.

Пусть $Y\in\mathscr{X}$ — некоторое векторное поле на M (не зависящее от λ). Тогда

$$\frac{d}{d\lambda}\int\left\langle Y,\ \mathscr{Y}(g,\ \pi)\right\rangle = \int\left\langle Y,\ \frac{d\mathscr{Y}(g,\ \pi)}{d\lambda}\right\rangle =$$
$$= \int\left\langle Y,\ \mathrm{D}\mathscr{Y}(g,\ \pi)\cdot\left(\frac{\partial g}{\partial\lambda},\ \frac{\partial\pi}{\partial\lambda}\right)\right\rangle =$$
$$= \int\left\langle Y,\ \mathrm{D}\mathscr{Y}(g,\ \pi)\cdot\left\{J\circ[\mathrm{D}\Phi(g,\ \pi)]^*\cdot\begin{pmatrix}N\\X\end{pmatrix}\right\}\right\rangle =$$

$(J^* = -J)$

$$= -\int\left\langle \mathrm{D}\Phi(g,\ \pi)\cdot\{J\circ[\mathrm{D}\mathscr{Y}(g,\ \pi)]^*\cdot Y\}\cdot\begin{pmatrix}N\\X\end{pmatrix}\right\rangle =$$
$$= -\int\left\langle \mathrm{D}\Phi(g,\ \pi)\cdot(L_Y g,\ L_Y\pi),\ (N,\ X)\right\rangle =$$

(инфинитезимальная ковариантность Φ)

$$= -\int\left\langle L_Y\Phi(g,\pi),\ (N,\ X)\right\rangle = -\int N L_Y\mathscr{H}(g,\pi) + \left\langle X, L_Y\mathscr{Y}(g,\pi)\right\rangle =$$

(интегрирование по частям)

$$= \int(L_Y N)\mathscr{H} + \int\left\langle L_Y X,\ \mathscr{Y}\right\rangle = \int Y(\mathrm{d}N)\mathscr{H} - \int\left\langle L_X Y, \mathscr{Y}\right\rangle = \int Y(\mathrm{d}N)\mathscr{H} +$$
$$+ \int\left\langle Y,\ L_X\mathscr{Y}\right\rangle.$$

Поскольку Y — произвольное поле,

$$\frac{d\mathscr{Y}}{d\lambda} = (dN)\mathscr{H} + L_X\mathscr{Y}. \quad ■$$

Теорему 6 можно переформулировать через скобки Пуассона, введенные в предыдущем разделе, следующим образом.

Теорема 7

Пусть даны $N_1, N_2 : M \to \mathbb{R}$, $X_1, X_2 : M \to TM$ *и*

$$F_1 = \int (N_1 \mathscr{H} + X_1 \cdot \mathscr{T}) \colon T^*\mathscr{M} \to \mathbb{R},$$

$$F_2 = \int (N_2 \mathscr{H} + X_2 \cdot \mathscr{Y}) \colon T^*\mathscr{M} \to \mathbb{R},$$

тогда

$$\{F_1, F_2\} = \int (L_{X_1} N_2 - L_{X_2} N_1) \mathscr{H} + \\ + \int \langle (N_1 \operatorname{grad} N_2 - N_2 \operatorname{grad} N_1), \mathscr{Y} \rangle + \langle L_{X_1} X_2, \mathscr{Y} \rangle,$$

и в частности

$$\left\{ \int N_1 \mathscr{H}, \int N_2 \mathscr{H} \right\} = \int \left(N_1 \operatorname{grad} N_2 - N_2 \operatorname{grad} N_1, \quad \mathscr{Y} \right),$$

$$\left\{ \int N \mathscr{H}, \int X \cdot \mathscr{Y} \right\} = - \int (L_X N) \mathscr{H},$$

$$\left\{ \int X_1 \cdot \mathscr{Y}, \int X_2 \cdot \mathscr{Y} \right\} = \int \langle L_{X_1} X_2 \cdot \mathscr{Y} \rangle.$$

Убедиться в том, что эти соотношения эквивалентны теореме 6, можно прямым вычислением. Мы будем называть их *каноническими коммутационными* соотношениями Дирака.

Для понимания расщепления, введенного Монкри [148], и конструкции, приводящей к пространству гравитационных степеней свободы (разд. 6), будет важна следующая инфинитезимальная версия теоремы 1.

Предложение 8

Пусть $(g, \pi) \in \mathscr{C}_{\mathscr{H}} \cap \mathscr{C}_\delta$. *Тогда*

$$\operatorname{im} \{ J \circ [D\Phi(g, \pi)]^* \} \subset \ker D\Phi(g, \pi).$$

Доказательство. Пусть $(h, \omega) \in \operatorname{im}\{J \circ [D\Phi(g, \pi)]^*\}$ и $(N, X) \in C^\infty \times \mathscr{X}$, такие, что $(h, \omega) = J \circ [D\Phi(g, \pi)^* \cdot (N, X)$. Пусть $(N(\lambda), X(\lambda))$ суть произвольные длительность и сдвиг, такие, что $(N(0), X(0)) = (N, X)$. Пусть $(g(\lambda), \pi(\lambda))$ — решение уравнения эволюции с длительностью и сдвигом $(N(\lambda), X(\lambda))$ и с начальными данными $(g, \pi) \in \mathscr{C}_{\mathscr{H}} \cap \mathscr{C}_\delta$. Поскольку $\Phi(g, \pi) = 0$, по теореме 6 имеем $\Phi(g(\lambda), \pi(\lambda)) = 0$ для всех λ, при которых существует решение. Отсюда

$$0 = \frac{d}{d\lambda} \Phi(g(\lambda), \pi(\lambda)) \Big|_{\lambda=0} = D\Phi(g(\lambda), \pi(\lambda)) \cdot \left(\frac{\partial g(\lambda)}{\partial \lambda}, \frac{\partial \pi(\lambda)}{\partial \lambda} \right) \Big|_{\lambda=0} =$$

$$= D\Phi(g(\lambda), \pi(\lambda)) \left\{ J \circ [D\Phi(g(\lambda), \pi(\lambda))]^* \cdot \begin{pmatrix} N(\lambda) \\ X(\lambda) \end{pmatrix} \right\} \Big|_{\lambda=0} =$$

$$= D\Phi(g, \pi) \cdot \left\{ J \circ [D\Phi(g, \pi)]^* \cdot \begin{pmatrix} N \\ X \end{pmatrix} \right\} = D\Phi(g, \pi) \cdot (h, \omega).$$

Поэтому $(h, \omega) \in \ker D\Phi(g, \pi)$. ■

Изучим теперь структуру многообразия множества связей $\mathscr{C}_{\mathscr{H}} \cap \mathscr{C}_\delta$. Наложим следующие условия на $(g, \pi) \in T^*\mathscr{M}$:

$C_{\mathscr{H}}$: Если $\pi=0$, то g — неплоская метрика.

C_δ: Если для $X \in \mathscr{X}(M)$ имеем $L_X g=0$ и $L_X \pi=0$, то $X=0$.

C_{tr}: $\mathrm{tr}\, \pi'$ постоянен на M.

Рассмотрим связи по отдельности, начиная с гамильтоновой связи.

Предложение 9

Пусть $(g, \pi) \in \mathscr{C}_{\mathscr{H}}$ удовлетворяет условию $\mathscr{C}_{\mathscr{H}}$. Тогда $\mathscr{C}_{\mathscr{H}}$ есть C^∞-подмногообразие многообразия $T^*\mathscr{M}$ в окрестности (g, π) с касательным пространством

$$T_{(g, \pi)}\mathscr{C}_{\mathscr{H}} = \ker D\mathscr{H}(g,\pi) .$$

Доказательство основано на ряде свойств эллиптических операторов и пространств Соболева. Мы коротко напомним относящиеся сюда факты (см. доказательства в работах [16, 157]).

Пусть Ω — ограниченная открытая область пространства $\mathbb{R}^n$ с гладкой границей. Для любой C^∞-функции f из $\mathbb{R}^n$ в $\mathbb{R}^m$ мы определим $W^{s,p}(\Omega, \mathbb{R}^m)$-норму f как

$$\|f\|_{W^{s,p}} = \sum_{0 \leqslant \alpha \leqslant s} \|D^\alpha f\|_{L_p(\Omega)},$$

где D^α— полная производная f порядка α и через $\|\ \|_{L_p(\Omega)}$ обозначена обычная L_p-норма на Ω:

$$\|g\|_{L_p(\Omega)} = \left(\int_\Omega |g(x)|^p\, dx\right)^{1/p} .$$

По определению $W^{s,p}(\Omega, \mathbb{R}^m)$ есть пополнение пространства $C^\infty(\Omega, \mathbb{R}^m) = \{$сужение C^∞-функций с $\mathbb{R}^n$ на $\Omega\}$ по этой норме. Будем употреблять более краткую запись $W^{s,p}$ для $W^{s,p}(\Omega, \mathbb{R}^m)$ и аналогичную для других подобных выражений в случаях, когда возможность путаницы будет невелика.

Для компактного многообразия M без края и векторного расслоения E над M будем обозначать через $W^{s,p}(E)$ пространство всех сечений E класса $W^{s,p}$ в некотором (и, значит, в каждом) покрытии M картами. Для функций с действительными значениями мы будем писать просто $W^{s,p}$, но в случаях других тензорных расслоений примем для $W^{s,p}(E)$ специальные обозначения, такие, как $\mathscr{M}^{s,p}$ для $W^{s,p}$-пространства римановых метрик.

В случае $p=2$ пространства $W^{s,p}$ обозначаются H^s. В этом, и только в этом случае мы имеем дело с гильбертовым пространством.

Предположим теперь, что имеются два векторных расслоения E и F над одним и тем же многообразием M и линейный дифференци-

альный оператор D порядка k:

$$D: C^\infty(E) \to C^\infty(F).$$

Линейный дифференциальный оператор порядка k есть отображение, такое, что для данных карт на E и F (и, следовательно, для всех карт) этот оператор принимает форму $D = \sum_{|\alpha| < k} a_\alpha(x) D^\alpha$, где $D^\alpha = \partial^{|\alpha|}/\partial x_1^{\alpha_1} \ldots \partial x_n^{\alpha_n}$ — частная производная в карте U многообразия M, $\alpha = (\alpha_1, \ldots, \alpha_n)$, $|\alpha| = \sum_{i=1}^n \alpha_i$ и $a_\alpha(x)$ есть линейная функция из модельного пространства для слоя E_x в модельном пространстве для слоя F_x над $x \in U$. На D можно смотреть как на отображение между пространствами Соболева:

$$D: W^{s+k,\,p} \to W^{s,\,p}.$$

Для оператора D есть L_2-сопряженный оператор D^*, определяемый, как обычно, уравнением

$$(Df,\ g)_{L_2} = (f,\ D^*g)_{L_2}, \text{ т. е. } \int_M \langle Df,\ g \rangle d\mu = \int_M \langle f,\ D^*g \rangle d\mu,$$

где $d\mu$ — некоторый предпочтительный элемент объема, скажем тот, который связан с метрикой: $d\mu(g) = [\det(g_{ij})]^{1/2} dx^1 \wedge \ldots \wedge dx^n$, и $\langle , \rangle$ означает внутреннее произведение в слоях. Однако надобность в этой структуре отпадает, если используется естественное сопряжение.

Дифференциальный оператор D является *эллиптическим*, если он обладает инъективным (главным) символом. Для каждого x в M и для каждого $\xi \in T_x^*M$ = (слой в кокасательном расслоении) *символ* $\sigma_\xi(D)$ есть линейное отображение из слоя E_x в слой F_x. По заданному выражению D в картах символ $\sigma_\xi(D)$ получается подстановкой компонент $\xi \in T_x^*M$ вместо соответствующих частных производных в членах с производными наибольшего порядка. Таким образом, $\sigma_\xi(x)$ для каждой координаты на F_x является однородным полиномом степени k по компонентам ξ. Например, символом для обычного лапласиана $\nabla^2 = \sum_{i=1} \partial^2/\partial x_i^2$ является $\sigma_\xi(\nabla^2) = \|\xi\|^2$.

Для эллиптических операторов имеет место следующая основная теорема расщепления.

Альтернатива Фредгольма: теорема 10

Если D *или* D^* — *эллиптический оператор, то* $W^{s,p}(F) = \mathrm{im}\ D \oplus \oplus \ker D^*$, *где сумма есть* L_2-*ортогональная прямая сумма.*

Доказательство предложения 9. Рассмотрим отображение $\mathcal{H}: T^*\mathcal{M} \to C_d^\infty$; $(g, \pi) \mapsto \mathcal{H}(g, \pi)$. Покажем, что при выполнении условия $C_{\mathcal{H}}$

$$D\mathcal{H}(g,\ \pi): T_{(g,\ \pi)}(T^*\mathcal{M}) = S_2 \times S_d^2 \to T_{\mathcal{H}(g,\ \pi)} C_d^\infty = C_d^\infty$$

есть сюръективное отображение с таким ядром расщепления, что

$\mathcal{H}$ является погружением в окрестности (g, π). Используя пространства Соболева и теорему о неявной функции и переходя затем к классу C^∞ через требование регулярности, мы получим, что $C_{\mathcal{H}} = \mathcal{H}^{-1}(0)$ является гладким подмногообразием в окрестности (g, π).

Из теоремы 10 следует сюрьективность $D\mathcal{H}(g, \pi)$ при условии, что его L_2-сопряжение

$$[D\mathcal{H}(g, \pi)]^*: C^\infty \longrightarrow S_d^2 \times S_2,$$

$$[D\mathcal{H}(g, \pi)]^* \cdot N = \Big\{ - NS_g(\pi, \pi) +$$

$$+ [N\operatorname{Ein}(g) - \operatorname{Hess} N - g\,\Delta N]^\# \, d\mu(g),\ 2N\left[(\pi')^\flat - \frac{1}{2}(\operatorname{tr}\pi')\,g\right]\Big\}$$

инъективно и обладает инъективным символом. Этот символ для $[D\mathcal{H}(g, \pi)]^*$ имеет вид

$$\sigma_\xi [D\mathcal{H}(g, \pi)]^* = [(-\xi \otimes \xi + g\|\xi\|^2)^\# \, d\mu(g),\ 0]:$$

$$\mathbb{R} \longrightarrow ((T_x^*M \otimes T_x^*M)_{\mathrm{sym}}\, d\mu(g),\ (T_xM \otimes T_xM)_{\mathrm{sym}})$$

для $\xi \in T_x^*M$. Для $s \in \mathbb{R}$, $\xi \neq 0$ из $(-\xi\otimes\xi + g\|\xi\|^2)s = 0$ после взятия следа следует, что $2\|\xi\|^2 s = 0$, а отсюда $s = 0$, т. е. символ инъективен.

Любое $N \in \ker[D\mathcal{H}(g, \pi)]^*$ удовлетворяет условиям

(I) $- NS_g(\pi, \pi) + [N\operatorname{Ein} g - \operatorname{Hess} N - g\,\Delta N]^\# \, d\mu(g) = 0,$

(II) $2N\left[(\pi') - \frac{1}{2}(\operatorname{tr}\pi')\,g\right] = 0.$

Взятие следа условия (II) дает $N(\operatorname{tr}\pi') = 0$, и потому из (II) вновь вытекает $N\pi = 0$. Таким образом, из условия (I) следует

(III) $N\operatorname{Ein}(g) - \operatorname{Hess} N - g\,\Delta N = 0$

След (III) дает

$$2\,\Delta N + \frac{1}{2} R(g)\,N = 0.$$

Однако из равенств $\mathcal{H}(g, \pi) = 0$ и $N\pi = 0$ вытекает $NR(g) = 0$. Отсюда

$$\Delta N = 0$$

и, следовательно, $N = \mathrm{const}$.

Если $\pi \neq 0$, то из $N\pi = 0$ следует, что $N = 0$, поскольку $N = \mathrm{const}$. Таким образом, отображение $[D\mathcal{H}(g,\pi)]^*$ инъективно, а потому $D\mathcal{H}(g, \pi)$ сюрьективно.

Если $\pi = 0$, то из (III) следует, что $N\operatorname{Ein}(g) = 0$, так как $N = \mathrm{const}$, и, следовательно, $N\operatorname{Ric}(g) = 0$. Следовательно, если $N \neq 0$, то $\operatorname{Ric}(g) = 0$, а потому g — плоская метрика, так как $\dim M = 3$. Но плоская метрика g и $\pi = 0$ исключаются условием $C_{\mathcal{H}}$. Следовательно, $N = 0$, и вновь отображение $D\mathcal{H}(g, \pi)$ сюрьективно. ■

Предложение 11

Если $(g, \pi) \in \mathscr{C}_\delta = \{(g, \pi) | \mathscr{J}(g, \pi) = 0\} \subset T^*\mathscr{M}$ *удовлетворяет условию* C_δ, *то* $\mathscr{C}_\delta$ *есть гладкое подмногообразие многообразия* $T^*\mathscr{M}$ *в некоторой окрестности* $\{g, \pi\}$ *с касательным пространством*

$$T_{(g,\ \pi)}\mathscr{C}_\delta = \ker [D\mathscr{J}(g,\ \pi)].$$

Доказательство. Оператор, сопряженный к производной $\mathscr{J}(g, \pi)$, определяется соотношением

$$[D\mathscr{J}(g,\ \pi)]^* \cdot X = (-L_X\pi,\ L_X g).$$

Символ инъективен (только из-за инъективности по второй компоненте). Ядром отображения $[D\mathscr{J}(g, \pi)]^*$ является $\{X / L_X\pi = 0, L_X g = 0\}$, так что инъективность $[D\mathscr{J}(g, \pi)]^*$ совпадает в точности с условием C_δ. Искомый результат вытекает тогда из теоремы о неявной функции, как в теореме 9. ■

Чтобы показать, что пересечение $\mathscr{C} = \mathscr{C}_{\mathscr{H}} \cap \mathscr{C}_\delta$ есть подмногообразие многообразия $T^*\mathscr{M}$, нужны дополнительные ограничения, поскольку могут быть точки, в которых это пересечение не трансверсально. В такой точке необходимо предположить, что (g, π) удовлетворяет условию $\operatorname{tr} \pi' = \text{const}$.

Теорема 12

Пусть $(g, \pi) \in \mathscr{C}_{\mathscr{H}} \cap \mathscr{C}_\delta$ *удовлетворяет условиям* $C_{\mathscr{H}}$, C_δ *и* C_{tr}; *тогда множество связей* $C = C_{\mathscr{H}} \cap C_\delta$ *есть* C^∞*-подмногообразие многообразия* $T^*\mathscr{M}$ *в окрестности «точки»* $\{g, \pi\}$ *с касательным пространством*

$$T_{(g,\ \pi)}\mathscr{C} = \ker D\Phi(g,\ \pi),$$

где $\Phi = \{\mathscr{H},\ \mathscr{J}\}$.

Доказательство. Нам нужно показать, что отображение $D\Phi(g, \pi) = (D\mathscr{H}(g, \pi),\ D\mathscr{J}(g, \pi))$ сюръективно, когда $(g, \pi) \in \mathscr{C}$, и удовлетворяет указанным условиям. Сопряженным к нему отображением является

$$[D\Phi(g,\ \pi)]^* : C^\infty \times \mathscr{X} \longrightarrow S_d^2 \times S_2,$$

$$(N,\ X) \mapsto [D\Phi(g,\ \pi)]^* \cdot (N,\ X) = [D\mathscr{H}(g,\ \pi)]^* \cdot N + [D\mathscr{J}(g,\ \pi)]^* \cdot X.$$

Для $\xi \in T_x^* M$, $\xi \neq 0$, как и прежде, можно показать, что символ этого отображения, $\sigma_\xi [D\Phi(g, \pi)]^* \cdot \xi \in T_x^* M$, инъективен (см., однако, замечания о различных типах эллиптичности в работе [189]). Таким образом, остается показать инъективность $[D\Phi(g, \pi)]^*$. Пусть $(N, X) \in \ker [D\Phi(g, \pi)]^*$. Тогда из формулы для $(D\Phi(g, \pi))^*$ имеем

$$\text{(I)}\quad -NS_g(\pi,\ \pi) + [N \operatorname{Ein}(g) - (\operatorname{Hess} N + g\,\Delta N)]^*\, d\mu(g) - L_X\pi = 0,$$

$$\text{(II)}\quad 2N\left[(\pi') - \frac{1}{2}(\operatorname{tr}\pi')\,g\right] + L_X(g) = 0.$$

Взяв след от (I) и (II), получим

(III) $-\frac{N}{2}\mathscr{H}(g, \pi)+2(\Delta N)\,d\mu(g)+\operatorname{tr} L_X\pi=0,$

(IV) $-N\operatorname{tr}\pi'+2\operatorname{div}X=0.$

Далее,

$$\operatorname{tr} L_X\pi = X\cdot d\operatorname{tr}\pi-\pi\cdot L_Xg+\operatorname{div}X(\operatorname{tr}\pi),$$

так как

$$L_X\pi=(L_X\pi')\otimes d\mu(g)+\pi'\otimes(\operatorname{div}X)\,d\mu(g)$$

в координатах

$$(L_X\pi)^{ij}=X^k\pi^{ij}_{|k}-\pi^{ik}X^j_{|k}-\pi^{jk}X^i_{|k}+X^k_{|k}\pi^{ij}.$$

Поскольку $\mathscr{H}(g, \pi)=0$, (III) сводится к равенству

(V) $2(\Delta N)\,d\mu(g)+X\cdot d\operatorname{tr}\pi-\pi\cdot L_Xg+(\operatorname{div}X)(\operatorname{tr}\pi)=0.$

Используя (II) и (IV), чтобы исключить L_Xg и $\operatorname{div}X$ в (V), получим

(VI) $2(\Delta N)+X\cdot d\operatorname{tr}\pi'-\pi'\cdot L_Xg+(\operatorname{div}X)(\operatorname{tr}\pi')=$

$$=2\Delta N+2N\pi'\cdot\left[(\pi')^\flat-\frac{1}{2}(\operatorname{tr}\pi')g\right]+\frac{N}{2}(\operatorname{tr}\pi')(\operatorname{tr}\pi')+X\cdot d\operatorname{tr}\pi'=$$

$$=2\,\Delta N+2N\pi'\cdot\pi'+\frac{N}{2}(\operatorname{tr}\pi')^2+X\cdot d\operatorname{tr}\pi'=$$

$$=2\,\Delta N+2N\left[\pi'\cdot\pi'-\frac{1}{4}(\operatorname{tr}\pi')^2\right]+X\cdot d\operatorname{tr}\pi'=0.$$

Теперь заметим, что коэффициент при N, а именно

$$P(\pi'\cdot\pi')=\pi'\cdot\pi'-\frac{1}{4}(\operatorname{tr}\pi')^2=\left[\pi'-\frac{1}{2}(\operatorname{tr}\pi')g\right]\cdot\left[\pi'-\frac{1}{2}(\operatorname{tr}\pi')g\right]$$

положительно определен. Таким образом, если $\operatorname{tr}\pi'$ постоянен, то (VI) принимает вид

$$2\,\Delta N+2P(\pi', \pi')N=0;$$

это означает, что $N=0$, если только $\pi'\neq 0$. Если же $\pi'=0$, то $N=\mathrm{const}$. В последнем случае из (I) следует, что $\operatorname{Ein}(g)=0$ и отсюда $\operatorname{Ric}(g)=0$, т. е. g — плоская метрика, поскольку $\dim M=3$. Однако случай $(g_F, 0)$, где g_F — плоская метрика, исключается условием $C_{\mathscr{H}}$. Поэтому $\pi'\neq 0$ и $N=0$. Тогда в силу (I) и (II) имеем $L_Xg=0$ и $L_X\pi=0$, что, по условию C_δ, означает $X=0$. Итак $(N, X)=(0, 0)$ и, следовательно $[D\Phi(g, \pi)]^*$ инъективно при выполнении условий $C_{\mathscr{H}}$, C_δ и C_{tr}. Тогда утверждение доказываемой теоремы следует из теоремы о неявной функции. ■

Замечание. Слабым местом нашего анализа является необходимость наложить условие $\operatorname{tr}\pi'=\mathrm{const}$, чтобы показать, что пересечение $\mathscr{C}_{\mathscr{H}}\cap\mathscr{C}_\delta$ есть многообразие. Возникает сомнение, не достаточно ли одних только условий $C_{\mathscr{H}}$ и C_δ, чтобы система (I) и (II)

была инъективна. Трудность, однако, в том, что в этой системе скажем (VI) и (II) для (N, X), член взаимодействия вида $X \cdot d\, \mathrm{tr}\, \pi'$, уже может оказаться препятствием для доказательства единственности для этой системы. Результаты Монкри, обсуждаемые в разд. 5, проливают свет на этот вопрос.

В работах [54, 141] обсуждается существование гиперповерхностей с постоянным $\mathrm{tr}\, \pi'$. На этих преимущественных гиперповерхностях и будут проверяться условия $C_{\mathscr{H}}$ и C_{δ}.

3. АБСТРАКТНАЯ ЗАДАЧА КОШИ И ГИПЕРБОЛИЧЕСКИЕ УРАВНЕНИЯ

В этом разделе дается сводка тех результатов общей теории гиперболических задач с начальными данными, которые понадобятся нам для общей теории относительности. Полные доказательства длинны и технически сложны, поэтому будут приведены лишь их идеи. За подробностями можно обратиться к работам [108, 112—114]. Приводимый здесь абстрактный подход предпочтителен тем, что в нем содержатся как частные случаи и симметричные системы первого порядка, так и гиперболические системы второго порядка и комбинации этих систем. Более того, этот подход дает наименее ограничительные результаты в смысле дифференцируемости.

Начнем с линейного случая с последующим переходом к нелинейному и приведем один полезный для дальнейшего результат относительно дифференцируемости отображения по времени t, затем покажем, как эти результаты применяются к гиперболическим системам.

Мы предполагаем, что читатель знаком с теорией линейных полугрупп (см., например, [106, 140, 181]).

Пусть X — банахово пространство и $G(X, M, \beta)$ означает множество генераторов A C_0-полугруппы $e^{tA} = U(t)$ на X, удовлетворяющих условию

$$\|U(t)\| \leqslant M e^{\beta}, \quad t \geqslant 0;$$

т. е., по теореме Хилла — Иосиды, $\lambda - A$ действует на X одно-однозначно и, кроме того,

$$\|(\lambda - A)^{-n}\| \leqslant \frac{M}{(\lambda - \beta)^n}, \quad n = 1, 2, \ldots, \lambda > \beta.$$

Если $M = 1$, мы будем говорить, что генератор A *квазиаккретивен* или что оператор $U(t)$ *квазиконтрактивен*. Именно этот класс линейных полугрупп важен для нас. Напомним, что для $\varphi \in D(A)$ ($D(A)$ — область определения A) $U(t)\varphi = \varphi(t)$ также лежит в $D(A)$ и удовлетворяет уравнению эволюции

$$\frac{d}{dt} \varphi(t) = A \varphi(t), \tag{1}$$

где $\varphi(\cdot)$ при вычислении производной по времени рассматривается как отображение $[0, \infty]$ на X.

Пусть X, Y — банаховы пространства, $Y \subset X$, причем включение непрерывно и плотно. Пусть $U(t, s)$ — семейство ограниченных операторов на X, определенное при $0 \leqslant s \leqslant t \leqslant T$; здесь $[0, T]$ — подходящим образом выбранный интервал времени; T может быть сколь угодно большим. Пусть $A(t)$ — семейство линейных генераторов на $X, Y \subset D(A(t))$, $0 \leqslant t \leqslant T$. Мы назовем $U(t, s)$ семейством (сильных) *эволюционных операторов* для A, если

(I) $U(s, s) = 1$ и $(t, s) \mapsto U(t, s)$ сильно непрерывно в X;

(II) $U(t, s) U(s, r) = U(t, r)$, $0 \leqslant r \leqslant s \leqslant t \leqslant T$;

(III) $U(t, s)$ есть ограниченный оператор из Y в Y и сильно непрерывен по (t, s);

(IV) $(\partial/\partial t) U(t, s)\varphi = A(t) U(t, s)\varphi$, $\varphi \in Y$ («дифференциальное уравнение вперед»), причем и правая, и левая части сильно непрерывны по (t, s) со значениями в $B(Y, X)$ (ограниченные операторы из Y в X) и $\partial/\partial t$ берется по норме пространства X.

Если продифференцировать (II) по s при $s = r$ и использовать (IV), то формально получим «дифференциальное уравнение назад»:

$$\frac{\partial}{\partial s} U(t, s)\varphi = -U(t, s) A(s)\varphi$$

для $\varphi \in Y$. Это легко доказать, представив (IV) в виде интегрального уравнения по времени; запишем

$$U(t, s)\varphi = \varphi + \int_s^t A(\tau) U(\tau, s)\, d\tau$$

и используем равенство

$$\frac{1}{h}[U(t, s+h)\varphi - U(t, s)\varphi] = \\ = U(t, s+h)\left[\frac{\varphi - U(s+h)\varphi}{h}\right] + \frac{1}{h}[U(s+h, s)\varphi - \varphi] - \\ - \frac{1}{h}\int_s^{s+h} A(\tau) U(\tau, s+h)\varphi\, d\tau.$$

Семейство $A(t) \in G(X, M, \beta)$ (при фиксированных M, β) называется *устойчивым*, если для любого $s_j \geqslant 0$ и $0 \leqslant t_1 \leqslant \ldots \leqslant t_k \leqslant T$

$$\exp[s_k A(t_k)] \exp[s_{k-1} A(t_{k-1})] \ldots \exp[s_1 A(t_1)] \leqslant \\ \leqslant M \exp[\beta(s_1 + \ldots + s_k)],$$

или, что то же самое,

$$\|(\lambda - A(t_k))^{-1} \ldots (\lambda - A(t_1))^{-1}\| \leqslant \frac{M}{(\lambda - \beta)^k}, \quad \lambda > \beta.$$

Если $A(t)\in G(X, 1, \beta)$, то $A(t)$ очевидным образом устойчиво. Если пространство X с некоторой новой нормой $\|\ \|_t$, зависящей от t экспоненциально

$$\|\varphi\|_t \leqslant \|\varphi\|_s e^{c|t-s|}, \quad s,\ t\in[0, T],$$

обозначить через X_t и если $A(t)\in G(X_t, 1, \beta)$, то $A(t)$ устойчиво в X_{t_0} с $M=e^{2cT\beta}$ (см. работу Като [110], предложение 3.4). Там же в предложении 3.5 показано, что ограниченное возмущение устойчивого семейства устойчиво.

В следующей теореме $L^\infty_*([0, T]; B(Y, X))$ означает класс эквивалентности *строго* измеримых существенно ограниченных функций из $(0, T)$ в $B(Y, X)$ и $\mathrm{Lip}_*([0, T]; B(Y, X))$ означает сильно неопределенные интегралы функций в $L^\infty_*([0, T]; B(Y, X))$.

Теорема 13 [111]

Предположим, что

(I) $A(t)$ — *устойчивое семейство генераторов в* X, $0\leqslant t\leqslant T$;

(II) $Y\subset X$ *с непрерывным плотным включением и* $D(A)\supset Y$;

(III) *имеется семейство* $S(t): Y\to X$ *изоморфизмов (на), таких, что*

$$S(t)\,A(t)\,S(t)^{-1}=A(t)+B(t),$$

где $B(t)\in B(X)$ — *ограниченный оператор на* X *и*

а) $t\longmapsto S(t)$ *лежит в* $\mathrm{Lip}_*([0, T]; B(Y, X))$;

б) $t\longmapsto B(t)$ *лежит в* $L^\infty_*([0, T]; B(X))$;

(IV) $t\longmapsto A(t)\in B(Y, X)$ *непрерывно по норме.*

Тогда для A *имеется единственное семейство сильных эволюционных операторов.* Доказательство см. в работе [111].

Случай, когда область определения $A(t)$ постоянна во времени, много проще. Здесь мы предположим, что $D(A(t))=Y$ и что $A\in\mathrm{Lip}_*([0, T]; B(Y, X))$. Тогда (III) будет выполнено при $B=0$ и

$$S(t)=\lambda-A(t), \quad \lambda>\beta.$$

Однако для тех гиперболических задач, которые мы хотим рассмотреть, области определения не обязательно постоянны. Случай постоянной области определения является предметом начальной работы Като [109]; см. также [181].

Неоднородную задачу

$$\frac{\partial u}{\partial t}=A(t)\,u+f(t), \quad u(0)=\varphi$$

можно исследовать с помощью следующего хитроумного приема Като [114]. Мы перепишем это уравнение как уравнение на $X\times\mathbb{R}$ и рассмотрим эквивалентную однородную проблему

$$\frac{\partial}{\partial t}\begin{pmatrix}u\\k\end{pmatrix}=\hat{A}(t)\begin{pmatrix}u\\k\end{pmatrix}, \quad u(0)=\varphi, \quad k(0)=1,$$

где

$$\hat{A}(t)=\begin{pmatrix} A(t) & f(t) \\ 0 & 0 \end{pmatrix}.$$

Тогда теорему 13 можно применить к $\hat{A}$.

Во многих нелинейных задачах часто удобно рассматривать присоединенные (зависящие от времени) линейные задачи Коши, и к ним также применима теорема 13.

Чтобы проиллюстрировать применение этой теоремы, рассмотрим два случая, которые более всего касаются нас, а именно: симметричные гиперболические системы первого порядка и гиперболические системы второго порядка. Мы будем рассматривать их на $\mathbb{R}^m$, но в силу гиперболичности этих уравнений полученные результаты могут быть локализованы и затем применены также и к компактным многообразиям (см. [104]).

Сначала рассмотрим симметричные гиперболические системы первого порядка по Фридрихсу [99] (см. также [84, 112—114]). Они имеют вид

$$a_0(t, x)\frac{\partial u}{\partial t}=\sum_{j=1}^{m} a_j(t, x)\frac{\partial u}{\partial x^j}+a(t, x)u, \tag{2}$$

где $u(t, x)\in\mathbb{R}^N$, a_j, a действительны. Сделаем следующие предположения:

(I) имеются постоянные матрицы a_j^∞, a^∞, такие, что

$$a_j-a_j^\infty,\ a-a^\infty\in C([0, T],\ H^0(\mathbb{R}^m))\cap L^\infty([0, T],\ H^s(\mathbb{R}^m)),$$
$$j=0,\ 1,\ \ldots,\ m,$$
$$a_0-a_0^\infty\in \mathrm{Lip}([0, T],\ H^{s-1}(\mathbb{R}^m));$$

здесь $H^s(\mathbb{R})^m$ — обычное пространство Соболева над $\mathbb{R}^m$ (пока без уточнения области определения) и $s>(m/2)+1$;

(II) a_j— симметричные матрицы;

(III) $a_0(t, x)\geqslant cI$ при некотором $c>0$.

Теорема 14

При этих условиях гипотезы теоремы 13 удовлетворяются, если принять

$$X=L^2(\mathbb{R}^m)=H^0(\mathbb{R}^m),$$
$$Y=H^s(\mathbb{R}^m),\quad 1\leqslant s'\leqslant s,$$
$$S(t)=(1-\Delta)^{s'/2},$$
$$A(t)=a_0(t,\ \cdot)^{-1}\left[\sum_{j=1}^{m} a_j(t,\ \cdot)\frac{\partial}{\partial x^j}+a(t,\ \cdot)\right]$$

(замыкание этого оператора на C_0^∞), т. е. уравнение (2) порождает

сильную эволюционную систему в L^2, *которая отображает* $H^{s'}$ *на* $H^{s'}$ *{регулярность}*.

Примечание. Областью определения оператора $A(t)$ не обязательно является $H^1(\mathbb{R}^m)$; например, a_j могут обращаться в нуль.

Идея доказательства состоит в следующем. Если введем на X энергетическую норму

$$\|\varphi\|_t^2 = \int_{\mathbb{R}^m} [a_0(t, x)\varphi(x)]\cdot\varphi(x)\,dx,$$

то обнаружим, что

$$A(t) \in G(X_t, 1, \beta),$$

причем

$$\beta = \sup_{t,\,x}\left| a(t, x) - \frac{1}{2}\sum_{j=1}^{m} \frac{\partial a_j}{\partial x^j}(t, x)\right|,$$

и величина β конечна в силу неравенств Соболева. Основная идея состоит теперь в том, чтобы получить оценку

$$\|[\lambda - A(t)]^{-1}\|_t \geqslant \frac{1}{\lambda - \beta},$$

которая в силу неравенства Шварца следует из неравенства

$$\langle[\lambda - A(t)]\varphi, \varphi\rangle_t \geqslant (\lambda - \beta)\|\varphi\|_t^2.$$

Последнее неравенство легко доказывается с использованием интегрирования по частям и симметрии a_j. Устойчивость $A(t)$ вытекает из того факта, что норма $\|\ \|_t$ зависит от t экспоненциально. Труднее всего доказать, что

$$B(t) = [S, A(t)]\,S^{-1},$$

(где $[\ ,\ ]$ — коммутатор) является ограниченным оператором на X. Этот коммутатор расписывается явно; решающим пунктом является оценка коммутатора

$$\left[S, \frac{\partial a_j}{\partial x^i}\right].$$

Искомые оценки этого коммутатора получаются с помощью длинной, но относительно прямой последовательности оценок типа Соболева. Подробности для $s'=1$ можно найти в работе [113], а общий случай вполне аналогичен этому.

Замечание. Результаты этого типа для выражения (2) приводились уже в давних работах Фридрихса [98] и Куранта и Гильберта [60]. Однако нигде не были отчетливо и точно сформулированы гипотезы о дифференцируемости, которые имеют решающее значение в нелинейных задачах. Попытка промежуточного характера была предпринята в работе [84], а затем эта формулировка была уточнена

и пояснена в работе Като [112]. Нынешняя единая формулировка, предложенная в работе [108], также принадлежит Като.

Теперь мы рассмотрим гиперболическую систему второго порядка. Она имеет вид

$$a_{00}(t, x)\frac{\partial^2 u}{\partial t^2}=\sum_{i,j=1}^{m} a_{ij}(t, x)\frac{\partial^2 u}{\partial x^i\,\partial x^j}+2\sum_{i=1}^{m} a_{0i}(t, x)\frac{\partial^2 u}{\partial t\,\partial x^i}+$$

$$+a_0(t, x)\frac{\partial u}{\partial t}+\sum_{i=1}^{m} a_i(t, x)\frac{\partial u}{\partial x^i}+a(t, x)\,u, \quad (3)$$

где опять $u(t, x)\in\mathbb{R}^N$, $a_{\alpha\beta}$, a_α, a — матричные функции $N\times N$, а также предполагается, что $s>(^1/_2)m+1$ и

(I) имеются постоянные матрицы $a^\infty_{\alpha\beta}$, a^∞_α, a^∞, такие, что

$$a_{\alpha\beta}-a^\infty_{\alpha\beta},\ a-a^\infty\in\mathrm{Lip}([0, T];\ H^{s-1}(\mathbb{R}^m))\subset L^\infty([0, T];\ H^s(\mathbb{R}^m));$$

(II) $a_{\alpha\beta}$ симметрична;

(III) $a_{00}(t, x)\geqslant cI$ для некоторого $c>0$;

(IV) сильная эллиптичность: имеется $\varepsilon>0$, так что

$$\sum_{i,j=1}^{M} a_{ij}(t, x)\,\xi_i\xi_j\geqslant\varepsilon\left(\sum_{j=1}^{m}\xi_i^2\right)$$

(матричное неравенство) для всех $\xi=(\xi_1, \ldots, \xi_m)\in\mathbb{R}^m$.

Теорема 15

При этих условиях предположения теоремы 13 справедливы, если принять, что

$$X=H^1(\mathbb{R}^m)\times H^0(\mathbb{R}^m),$$
$$Y=H^{s'+1}(\mathbb{R}^m)\times H^{s'}(\mathbb{R}^m),\quad 1\leqslant s'\leqslant s,$$
$$S=(1-\Delta)^{s'/2}\times(1-\Delta)^{s'/2},$$

$$A(t)=\begin{pmatrix} 0 & I \\ a_{00}^{-1}\left[\sum a_{ij}\frac{\partial^2}{\partial x^i\,\partial x^j}+\sum a_j\frac{\partial}{\partial x^j}+a\right] & a_{00}^{-1}\left[2\sum a_{0j}\frac{\partial}{\partial x^j}+a_0\right]\end{pmatrix}$$

(*замыкание этого оператора на* C_0^∞,), *т. е. уравнение* (3) *порождает сильную эволюционную систему в* X, *которая отображает* Y *в* Y.

Здесь мы записали (3) обычным образом, как систему для $(u, \dot u)$ (первого порядка по времени).

Используем норму

$$\|(\varphi, \dot\varphi)\|_t^2=\int_{\mathbb{R}^m}\left[\sum_{i,j=1}^{m} a_{ij}(t, x)\frac{\partial\varphi}{\partial x^i}\cdot\frac{\partial\varphi}{\partial x^j}+c\varphi\cdot\varphi+a_{00}(t, x)\,\dot\varphi\cdot\dot\varphi\right]dx,$$

где постоянная c выбирается достаточно большой. Ввиду неравенства Гординга это дает эквивалентную норму на X (если воспользоваться сильной эллиптичностью). Тогда прямыми вычислениями получаем оценку

$$\|[\lambda - A(tS)]^{-1}\|_t \geqslant \frac{1}{\lambda-\beta},$$

для чего, как и прежде, нужно показать, что

$$\langle[\lambda - A(t)]u, u\rangle_t \geqslant (\lambda-\beta)\|u\|_t^2.$$

Можно также показать, как это сделано в работе [181], что $\lambda - A(t)$ есть одно-однозначное отображение на X, поэтому $A(t) \in G(X_t, 1, \beta)$, причем, как и выше, $\|\cdot\|_t$ зависит от t экспоненциально и, следовательно, $A(t)$ есть устойчивое семейство.

И вновь доказательство ограниченности $B(t)$ требует получения оценок для коммутаторов; подробности см. в работе [108].

Для дальнейшего использования в нелинейной задаче предположений о дифференцируемости (и в лемме 22) очень важно, чтобы они были именно такими, как они здесь сформулированы.

Замечание. Ясно, что аналогичным образом можно трактовать комбинированные системы таких уравнений, поскольку абстрактная теорема 13 включает как частные случаи системы (2) и (3). Это существенно для некоторых видов взаимодействия материальных полей с гравитационным полем.

Теперь обратимся к нелинейной задаче. Как и выше, пусть X и Y — банаховы пространства, причем Y плотно и непрерывно содержится в X. Пусть $W \subset Y$ есть открытая область, $T > 0$ и пусть $G: [0, T] \times W \to X$ — некоторое данное отображение. Нелинейное эволюционное уравнение имеет вид

$$\dot{u}(t) = G(t, u(t)), \quad \text{где} \quad \dot{u} = \frac{du}{dt}. \tag{4}$$

Если даны $s \in [0, T]$ и $\phi \in W$, то кривая-решение (или интегральная кривая) для G со значением ϕ в s есть отображение $u(\cdot) \in C^0([s, T], W) \cap ([s, T], X)$, такое, что на $[s, T]$ выполняется (4) и $u(s) = \phi$.

Если эти кривые-решения существуют и единственны для ϕ в открытом множестве $U \subset W$, мы можем ввести эволюционные операторы $F_{t,s}$: $U \to W$, которые отображают $u(s) = \phi$ на $u(t)$. Будем говорить, что система (4) *корректна* (или *коши-устойчива*), если оператор $F_{t,s}$ непрерывен (по Y-топологии на U и W) при любых s, t, удовлетворяющих условию $0 \leqslant s \leqslant t \leqslant T$. Заметим, что непрерывность $F_{t,s}(\phi)$ по (t, z, ϕ) вместе вытекает из общих гипотез [30]. Более того, если имеется корректность для малых временны́х интервалов, легко доказать ее для максимально расширенного потока.

Установить корректность в конкретных, особенно «гиперболи-

ческих», случаях может оказаться затруднительным. Непрерывность оператора $F_{t,s}$ из Y в Y, вообще говоря, нельзя заменить более сильными условиями гладкости, такими, как непрерывность по Липшицу или даже по Гёльдеру; Като [112] дал простой пример, показывающий это, а именно уравнение $\dot{u}+uu_x=0$ в $Y=H^{s+1}$, $X=H^s$ на $\mathbb{R}$. Эти вопросы гладкости будут обсуждаться ниже.

Наиболее строго изучены те нелинейные эволюционные уравнения, которые приводят к нелинейным сжимающим полугруппам, порожденным монотонными операторами [20]. Их эволюционные операторы иногда бывают определены на всем пространстве X. Это нетипично для гиперболических задач, где оператор $F_{t,s}$ может быть определенным лишь на Y, непрерывным при отображении из Y в Y, дифференцируемым при отображении Y в X и Y-локально липшицевым при отображении из X в X, не будучи X-локально липшицевым при отображении из X в X или Y-локально липшицевым при отображении из Y в Y; это легко увидеть в приведенном выше примере.

Конкретизируя (4), рассмотрим квазилинейную абстрактную задачу Коши

$$\dot{u}=A(t, u)u+f(t, u), \quad 0\leqslant t\leqslant T, \quad u(0)=\varphi, \tag{5}$$

где u принимает значения в X и $A(t, u)$ есть (неограниченный) линейный оператор, зависящий от неизвестной u нелинейным образом. Мы включили сюда для полноты неоднородный член f, хотя он может быть устранен упомянутым выше приемом Като.

Сделаем следующие допущения. Будем исходить из того, что есть четыре (действительных) банаховых пространства

$$Y\subset X\subset Z'\subset Z,$$

причем все они рефлексивны и сепарабельны, а включение непрерывно и плотно. Предположим, что

$(Z')Z'$ есть интерполирующее пространство между Y и Z; таким образом, если $U\in B(Y)\cap B(Z)$, то $U\in B(Z')$, причем $\|U\|_{Z'}\leqslant c\max\{\|U\|_Y, \|U\|_Z\}$; $B(Y)$ означает множество ограниченных операторов на Y.

Пусть $N(Z)$ есть множество всех норм в Z, эквивалентных некоторой данной $\|\ \|_Z$. Тогда $N(Z)$ будет метрическим пространством с функцией расстояния

$$d(\|\ \|_\mu, \|\ \|_\nu)=\ln\max\left\{\sup_{0\neq z\in Z}\|z\|_\mu/\|z\|_\nu, \sup_{0\neq z\in Z}\|z\|_\nu/\|z\|_\mu\right\}.$$

Введем теперь на $[0, T]\times W$, где $T>0$ и W — открытое множество в Y, четыре функции A, N, S, f со следующими свойствами.

Для всех $t, t', \ldots, \in[0, T]$ при всех $w, w', \ldots, \in W$ найдутся действительное число β и положительные числа $\lambda_N, \mu_N, \ldots$, такие, что будут выполнены следующие условия:

$(N) N(t, w) \in N(Z)$, причем

$$d(N(t, w), \| \|_Z) \leqslant \lambda_N,$$
$$d(N(t', w'), N(t, w)) \leqslant \mu_N(|t-t'| + \|w'-w\|_X).$$

(S) $S(t, w)$ есть изоморфизм Y на Z, причем

$$\|S(t, w)\|_{Y,Z} \leqslant \lambda_S, \quad \|S(t, w)^{-1}\|_{Z,Y} \leqslant \lambda'_S,$$
$$\|S(t', w') - S(t, w)\|_{Y,Z} \leqslant \mu_S(|t'-t| + \|w'-w\|_X).$$

(A1) $A(t, w) \in G(Z_{N(t,w)}, 1, \beta)$, где $Z_{N(t,w)}$ означает банахово пространство Z с нормой $N(t, w)$. Это значит, что $A(t, w)$ есть C_0-генератор в Z, такой, что $\|e^{\tau A(t,w)}z\| \leqslant e^{\beta\tau}\|z\|$ для всех $\tau \geqslant 0$ и $z \in Z$.

(A2) $S(t, w) A(t, w) S(t, w)^{-1} = A(t, w) + B(t, w)$, $B(t, w) \in B(Z)$, $\|B(t, w)\|_Z \leqslant \lambda_B$.

(A3) $A(t, w) \in B(Y, X)$, причем $\|A(t, w)\|_{Y,X} \leqslant \lambda_A$, $\|A(t, w') - A(t, w)\|_{Y,Z'} \leqslant \mu_A \|w'-w\|_{Z'}$, и отображение $t \mapsto A(t, w) \in B(Y, Z)$ непрерывно по норме.

(f1) $f(t, w) \in Y$, $\|f(t, w)\|_Y \leqslant \lambda_f$, $\|f(t, w') - f(t, w)\|_{Z'} \leqslant \mu_f \|w'-w\|_Z$, и $t \mapsto f(t, w) \in Z$ непрерывно.

Замечания. (I) Если $N(t, w) = \text{const} = \| \|_Z$, условие (N) излишне. Если $S(t, w) = \text{const} = S$, условие (S) тривиально. Если имеет место и то, и другое и, кроме того, $X = Z' = Z$, мы приходим к случаю, рассмотренному в [113].

(II) В большинстве приложений можно выбрать $Z' = Z$ и (или) $Z' = X$

(III) В статье [108] имелось дополнительное условие (A4), которое, как показал затем Като [114], является излишним.

Теорема 16

Допустим, что удовлетворяются условия (Z′), (N), (S), (A1) — A(3) *и* (f1). *Тогда существуют такие положительные постоянные* ρ' *и* $T' \leqslant T$, *что уравнение* (5) *имеет единственное на* $[0, T']$ *решение* u, *если* $\phi \in Y$ *и* $\|\phi - y_0\|_Y \leqslant \rho'$, *причем*

$$u \in C^0([0, T']; W) \cap C^1([0, T']; X).$$

Здесь ρ' зависит только от $\lambda_N, \lambda_S, \lambda'_S$ и $R = \operatorname{dist}(y_0, Y \setminus W)$, в то время как T' может зависеть от всех постоянных $\beta, \lambda_N, \mu_N, \ldots$ и R. Когда ϕ изменяется в Y с выполнением условия $\|\phi - y_0\|_Y \leqslant \rho'$, отображение $\phi \mapsto u(t)$ липшиц-непрерывно по Z'-норме равномерно по $t \in [0, T']$.

Для установления корректности мы должны усилить некоторые предположения. Наложим еще следующие условия:

$$\text{(B)} \quad \|B(t, w') - B(t, w)\|_Z \leqslant \mu_B \|w'-w\|_Y,$$
$$\text{(f2)} \quad \|f(t, w') - f(t, w)\|_Y \leqslant \mu'_f \|w'-w\|_Y.$$

Теорема 17

Допустим, что удовлетворяются условия (Z′), (N), (S), (A1) — (A3), (B), f(1), f(2) *и в них* $S(t, w)$ *предполагается независимым от* w. *Тогда существует положительная постоянная* $T'' \leqslant T'$, *такая, что когда* ϕ *пробегает* Y *с выполнением условия* $\|\phi - y_0\|_Y \leqslant \rho'$, *отображение* $\phi \mapsto u(t)$, *определяемое теоремой* 16, *непрерывно по норме* Y *равномерно по* $t \in [0, T'']$.

Замечание. Как и в работе [113], можно доказать аналогичную теорему непрерывности для случая, когда варьируются не только начальное значение ϕ, но и функции N, A, f, т. е. решение «устойчиво» при изменении самих уравнений. Наоборот, изменение S, по-видимому, довольно трудно рассмотреть.

Данная теорема, таким образом, гарантирует существование (заданных локально) отображений

$$F_{t,s}: Y \to Y,$$

которые непрерывны по всем переменным. Как и в линейном случае, имеем

$$F_{t,s} = \mathrm{Id},$$
$$F_{t,s} \circ F_{s,r} = F_{t,r}.$$

Мы будем называть $F_{t,s}$ *эволюционными операторами, порожденными уравнением* (5). Общее понятие эволюционного оператора для (4) вводится по аналогии.

Идея, на которой основано доказательство теоремы 17, состоит в том, чтобы фиксировать кривую $v(t)$, $v(0) = \phi$ в Y и рассматривать $u(t)$ как решение «задачи с замороженными коэффициентами»

$$\dot{u} = A(t, v)\,u + f(t, v), \quad u(0) = \phi,$$

что допустимо в связи с теоремой 13. Это приводит к отображению Φ: $v \mapsto u$, и мы ищем фиксированную точку Φ. В подходящем функциональном пространстве и для достаточно малого T' Φ будет в действительности сжатием и потому будет иметь единственную фиксированную точку.

Однако доказать, что u непрерывно зависит от ϕ, не так просто, и необходимы детальные оценки по линейной теории. Доказательство неизбежно должно оказаться более или менее затруднительным, поскольку зависимость от ϕ, вообще говоря, не является локально-липшицевой. За деталями доказательств читатель отсылается к работам [108, 113, 114].

Непрерывная зависимость решения от ϕ естественным образом приводит нас к вопросу о том, является ли эта зависимость в каком-нибудь смысле гладкой. Это важно для изучения соотношения между нелинейными теориями и их линеаризациями. Следующие результаты взяты нами из нескольких неопубликованных заметок Дорро и Марсдена.

Сначала введем подходящее понятие дифференцируемости для генератора G уравнения (4). Пусть X и Y — банаховы пространства, причем включение $Y \subset X$ непрерывно и плотно. Пусть $U \subset Y$ — открытое множество и f: $U \to X$ — некоторое заданное отображение. Будем говорить, что f α-дифференцируемо, если для каждого $x \in U$ найдется ограниченный линейный оператор $\mathrm{D}f(x)$: $Y \to X$, такой, что

$$\frac{\|f(x+h)-f(x)-\mathrm{D}f(x)\cdot h\|_X}{\|h\|_X} \to 0,$$

когда $\|h\|_Y \to 0$. Если f является α-дифференцируемым и $x \mapsto \mathrm{D}f(x) \in B(Y, X)$ непрерывно по норме, мы назовем отображение f C^{-1} — *α-дифференцируемым*. Отметим, что это свойство сильнее, чем дифференцируемость в смысле Фреше. Если отображение f α-дифференцируемо и отношение

$$\|f(x+h)-f(x)-\mathrm{D}f(x)\cdot h\|_X/\|h\|_X$$

равномерно ограниченно для x и $x+h$ в некоторой T-окрестности каждой точки, мы говорим, что f *локально-равномерно α-дифференцируемо.*

В наиболее часто встречающихся случаях α-дифференцируемость может быть установлена с помощью следующего предложения.

Предложение 18

Допустим, что $f: U \subset Y \to X$ — отображение класса C^2, и локально по топологии Y функция

$$x \mapsto \frac{\|\mathrm{D}^2 f(x)(h, h)\|_X}{\|h\|_Y \|h\|_X}$$

ограниченна. Тогда f локально-равномерно C^1 — α-дифференцируемо.

Это утверждение легко доказать, исходя из тождества

$$f(x+h)-f(x)-\mathrm{D}f(x)\cdot h = \int_0^1 \int_0^1 \mathrm{D}^2 f(x+sth)\, ds\, dt.$$

Теперь мы обратимся к соответствующему понятию для эволюционных операторов.

Отображение $g: U \subset Y \to X$ называется *β-дифференцируемым*, если оно α-дифференцируемо и $\mathrm{D}g(x)$ для каждого $x \in U$ расширяется до ограниченного оператора из X в X.

β-дифференцируемые отображения подчиняются правилу цепи. Например, если $g_1: Y \to Y$, $g_2: Y \to Y$ и каждое из них β-дифференцируемо (как отображение из Y в X) и непрерывно из Y в Y, то $g_2 \circ g_1$ β-дифференцируемо, причем, конечно,

$$D(g_2 \circ g_1)(x) = Dg_2[g_1(x)] \circ Dg_1(x).$$

Доказательство этого факта стандартно. В частности, можно применить правило цепи к $F_{t,s} \circ F_{s,r} = F_{t,r}$, если каждое из $F_{t,s}$ β-дифференцируемо. Дифференцирование этого равенства по s при $s=r$ дает «уравнение назад» для $x \in Y$:

$$\frac{\partial}{\partial s} F_{t,s}(x) = -\mathrm{D}F_{t,s}(x) \cdot G(x).$$

Затем дифференцирование по r при $r=s$ дает

$$\mathrm{D}F_{t,s}(x) \cdot G(x) = G[F_{t,s}(x)],$$

т. е. инвариантность потока генератора.

Получение строгих доказательств этих утверждений мы предоставляем читателю, который должен руководствоваться ходом рассуждений в линейном случае.

Для следующей теоремы мы сделаем такие предположения: $Y \subset X$ есть непрерывное и плотное включение, $F_{t,s}$ — непрерывная эволюционная система на открытом подмножестве $D \subset Y$, X-инфинитезимальный генератор $G(t)$ системы $F_{t,s}$ имеет областью определения D[1]). Кроме того, предположим:

(H_1) $G(t) : D \subset Y \to X$ локально-равномерно C^1-α-дифференцируем. Его производная обозначается $\mathrm{D}_x G(t, x)$ и предполагается, что она сильно непрерывна по t.

(H_2) Для $x \in D$; $s \geqslant 0$ пусть $T_{x,s}$ будет время жизни x вне s, т. е. задано $\sup \{t \geqslant s | F_{t,s}(x)\}$. Предположим, что в X имеется сильно непрерывная линейная эволюционная система $\{U^{x,s}(\tau, \sigma): 0 \leqslant \sigma \leqslant \tau \leqslant T_{x,s}\}$ в X, у которой X-инфинитезимальный генератор является расширением $\{\mathrm{D}_x G(t, F_{t,s}x) \in B(Y, X); 0 \leqslant t \leqslant T_{x,s}\}$, т. е., если $y \in Y$,

$$\frac{\partial}{\partial \tau^t} U^{x,s}(\tau,\sigma) \cdot y|_{\tau=0} = \mathrm{D}_x G(\tau, F_{t,s}(x)) \cdot y.$$

Теорема 19 (Дж. Дорро)

При сформулированных выше предположениях $F_{t,s}$ α-дифференцируема в точке x и

$$\mathrm{D}F_{t,s}(x) = U_{x,s}(t, s).$$

Доказательство. Введем $\varphi_t(x, y) = \varphi(t, x, y)$ с помощью равенства

$$G(t, x) - G(t, y) = \mathrm{D}_x G(t, y) \cdot (x - y) + \| x - y \|_X \varphi_t(x, y)$$

(или равно нулю, если $x=y$) и заметим, что в силу локальной равномерности $\|\varphi(t, x, y\|_X$ равномерно ограничены, если x и y близки по топологии Y. По совместной непрерывности $F_{t,s}(x)$, для $0<t<$

[1]) Как и в линейном случае, $G(t)$ может иметь расширение на бо́льшую область, но здесь нас интересует только $G(t)$ на D.

$<T_{xs}$ норма $\|\varphi(t, F_{t,s}(y), F_{t,s}(x)\|_X$ ограничена при $0\leqslant s\leqslant T$ при достаточно малых $\|x-y\|_Y$.

По построению имеем уравнение

$$\frac{d}{dt}F_{t,s}(x)=G[F_{t,s}(x)],\ 0\leqslant s\leqslant t\leqslant T_{x,s},\quad x\in D.$$

Пусть

$$w(t, s)=F_{t,s}(y)-F_{t,s}(x),$$

так что

$$\frac{\partial w(t, s)}{\partial t}=G(t, F_{t,s}(y))-G(t, F_{t,s}(x))=$$
$$=\mathrm{D}_xG(t, F_{t,s}(x))\,w(t, s)+\|w(t, s)\|_X\,\varphi(t, F_{t,s}(y)\,F_{t,s}(x)).$$

Поскольку $\mathrm{D}_xG(t, F_{t,s}(x))\cdot w(t, s)$ непрерывно по t, s со значениями в X, то, обозначая $U=U_{x,x}$, получим «дифференциальное уравнение назад»:

$$\frac{\partial}{\partial\sigma}U(t,\sigma)\,w(\sigma, s)=U(t,\sigma)\frac{\partial w(\sigma, s)}{\partial\sigma}-U(t,\sigma)D_xG(\sigma, F_{\sigma,s}(x))\cdot w(\sigma, s)=$$
$$=U(t,\sigma)\,\|w(\sigma, s)\|_X\,\varphi(\sigma, F_{\sigma,s}(y), F_{\sigma,s}(x)).$$

Отсюда, интегрируя от $\sigma=s$ до $\sigma=t$, получим

$$w(t, s)=U(t,s)(y-x)+\int_s^t U(t,\sigma)\,\|w(\sigma, s)\|_X\,\varphi(\sigma, F_{\sigma,s}(y), F_{\sigma,s}(x))\,d\sigma.$$

Пусть $\|U(\tau, \sigma)\|_{X,X}\leqslant M$ и $\|\varphi[\sigma, F_{\sigma,s}(y), F_{\sigma,s}(x)\|_X\leqslant M_2$, $0\leqslant s\leqslant\sigma\leqslant\leqslant\tau\leqslant T$.

Тогда из неравенства Гронволла имеем

$$\|w(t, s)\|_x\leqslant M_1e^{M_1M_2T}\|y-x\|_X=M_3\|y-x\|_X.$$

Иначе говоря,

$$\frac{\|F_{t,s}(y)-F_{t,s}(x)-U(t, s)(y-x)\|_X}{\|y-x\|_X}\leqslant$$
$$\leqslant M_1M_3\int_s^t\|\varphi(\sigma, F_{\sigma,s}(y), F_{\sigma,s}(x))\|_X\,d\sigma.$$

Из теоремы об ограниченной сходимости мы заключаем, что оператор $F_{t,s}$ β-дифференцируем в точке x и что $\mathrm{D}F_{t,s}(x)=U(t, s)$. (Функция $\varphi(t, F_{t,s}(y), F_{t,s}(x))$ сильно измерима в s, поскольку $\varphi(x, y)$ непрерывна при $x\neq y$.) ■

Этим завершается наше описание абстрактной нелинейной теории. Теперь мы установим, как эта нелинейная теорема существования и единственности применяется к квазилинейным уравнениям типа уравнений (2) и (3).

Сначала рассмотрим случай уравнения первого порядка

$$a_0(t, x, u)\frac{\partial u}{\partial t}=\sum_{j=1}^{m} a_j(t, x, u)\frac{\partial u}{\partial x^j}+a(t, x, u). \tag{6}$$

Предположим, что

(I) $s>\frac{1}{2}m+1$ и a_α, a — функции класса C^{s+1} по переменным t, x, u (и, возможно, по u определены локально);

(II) по u локально-равномерно выполняются линейные условия (I)—(III) теоремы 14.

Теорема 20

При этих условиях для уравнения (6) *справедливы теоремы* 16, 17 *и* 19, *т. е.* (6) *порождает единственную локальную эволюционную систему* $F_{t,s}$ *в* $X=H^{s-1}(\mathbb{R}^m)$ *при* $Y=H^s(\mathbb{R}^m)$ *и* $Z=Z'=L^2(\mathbb{R}^m)$. $F_{t,s}$ *отображает* Y *в* Y *непрерывно и при фиксированных* t, s *β-дифференцируемо как отображение* Y *в* X.

Доказательство со всеми подробностями требует длинных рассуждений об оценках в пространстве Соболева для проверки выполнения принятых гипотез, но оно довольно прямое. Подробности см. в работах [112, 113]. Заметим, что можно также выбрать $X=Z=Z'=L^2(\mathbb{R}^m)$, $Y=H^S(\mathbb{R}^m)$, но выбор, сделанный в теореме 20, подходит для теоремы 19. И опять громоздкость и технический характер деталей применения теоремы 19 к уравнению (6) удерживает нас от их изложения. Это снова полустандартное упражнение по теории пространства Соболева.

В случае уравнения второго порядка мы поступим следующим образом. Рассмотрим

$$a_{00}(t, s, u, \nabla u)\frac{\partial^2 u}{\partial t^2}=\sum_{i,j=1} a_{ij}(t, x, u, \nabla u)\frac{\partial^2 u}{\partial x^i\,\partial x^j}+$$
$$+2\sum_{j=1}^{m} a_{0j}(t, x, u, \nabla u)\frac{\partial^2 u}{\partial t\,\partial x^j}+a(t, x, u, \nabla u). \tag{7}$$

Здесь

$$\nabla u=\left(\frac{\partial u}{\partial x^1}, \ldots, \frac{\partial u}{\partial x^m}, \frac{\partial u}{\partial t}\right).$$

Предположим, что

(I) $a_{\alpha\beta}$, a — функции класса C^{s+1} по всем переменным (и, возможно, по u определены локально);

(II) по u локально-равномерно выполняются линейные условия (I)—(III) теоремы 15.

Теорема 21

I) *Если* $s>\frac{1}{2}m+1$, *то для уравнения* (7) *справедливы теоремы* 16, 17 *и* 19, *причем*

$$X=H^s(\mathbb{R}^m)\times H^{s-1}(\mathbb{R}^m),$$
$$Z=Z'=H^1(\mathbb{R}^m)\times H^0(\mathbb{R}^m),$$
$$Y=H^{s+1}(\mathbb{R}^m)\times H^s(\mathbb{R}^m),$$

т. е. (4.7) *порождает единственную локальную эволюционную систему* $F_{t,s}: Y\to Y$, *которая непрерывна и при фиксированных* t, s β-*дифференцируема как отображение из* Y *в* X.

(II) *Если* $a_{\alpha\beta}$ *не зависит от* ∇u, *те же утверждения справедливы при* $s>{}^1/_2\,m$.

Детали доказательства см. в работе [108].

Как будет видно из следующего раздела, с общей теорией относительности связан случай (II). Заметим также, что при $m=3$ решения $(u, \dot{u})$ будут лежать в $Y=H^r\times H^{r-1}$, где $r>2{,}5$. Например, в этом случае уравнение (7) задает корректную задачу для u в H^3. (Отметим, что при этом u принадлежит лишь к классу C^1 и нет необходимости, чтобы u принадлежало классу C^2.) Для гиперболических систем теоремы 20 и 21 являются наиболее сильными из известных результатов, хотя эти задачи рассматривались многими авторами [1]), такими, как Шоке-Брюа [32, 35], Курант и Гильберт [60], Дион [67], Френкл [96], Кржижанский и Шаудер [118], Лере [128], Лихнерович [133], Лайонз [137], Петровский [159], Шаудер [164] и Соболев [169].

4. ЗАДАЧА КОШИ ДЛЯ ТЕОРИИ ОТНОСИТЕЛЬНОСТИ

Мы начнем с задачи в пустоте, а затем перейдем к рассмотрению гравитации, взаимодействующей с другими полями. Дадим вначале обзор классических работ Лихнеровича [130] и Шоке-Брюа [32] и рассмотрим введение гармонических координат. Мы будем кратки, поскольку все это изложено в работах [35, 104]. Наш главный результат состоит в том, что для пространства-времени класса H^s с $s>2{,}5$ имеют место достаточно интересные теорема существования 23 и теорема единственности 27 для решения задачи Коши. Этот результат является в настоящее время наиболее сильным из известных утверждений для задачи Коши.

Для простоты примем, что мы имеем дело с $\mathbb{R}^4$, но ввиду гиперболичности это не ведет к существенной потере общности. В релятивистской теории пустого пространства мы ищем лоренцеву метрику $g_{\mu\nu}(t, x^i)$, при которой кривизна Риччи $R_{\mu\nu}$ равна нулю, т. е.

---

[1]) В теории относительности некоторые частные результаты для H^3 были приведены Хокингом и Эллисом ([104], с. 251).

$g_{\mu\nu}(t, x^i)$ должна удовлетворять системе

$$\left((R_{\mu\nu}\, t, x^i, g_{\mu\nu}, \frac{\partial g_{\mu\nu}}{\partial x^\alpha}, \frac{\partial^2 g_{\mu\nu}}{\partial x^\alpha \partial x^\beta}\right) = -\frac{1}{2} g^{\alpha\beta} \frac{\partial^2 g_{\mu\nu}}{\partial x^\alpha \partial x^\beta} -$$
$$-\frac{1}{2} g^{\alpha\beta} \frac{\partial^2 g_{\alpha\beta}}{\partial x^\mu \partial x^\nu} + \frac{1}{2} g^{\alpha\beta} \frac{\partial^2 g_{\alpha\nu}}{\partial x^\beta \partial x^\mu} + \frac{1}{2} g^{\alpha\beta} \frac{\partial^2 g_{\alpha\mu}}{\partial x^\beta \partial x^\nu} + H_{\mu\nu}\left(g_{\mu\nu}, \frac{\partial g_{\mu\nu}}{\partial x^\alpha}\right) = 0,$$

где $H_{\mu\nu}(g_{\mu\nu}, \partial g_{\mu\nu}/\partial x^\alpha)$ представляет собой рациональную комбинацию из $g_{\mu\nu}$ и $\partial g_{\mu\nu}/\partial x^\alpha$ со знаменателем $\det g_{\mu\nu} \neq 0$. Отметим, что контравариантный тензор $g_{\mu\nu}$ является рациональной комбинацией из $g_{\mu\nu}$ со знаменателем $\det g_{\mu\nu} \neq 0$.

Введем $G_{\mu\nu} = R_{\mu\nu} - {}^1/_2 g_{\mu\nu} R$ — тензор Эйнштейна, где $R = g^{\alpha\beta} R_{\alpha\beta}$ — скалярная кривизна. Тогда, как известно, G^0_μ содержит производные $g_{\mu\nu}$ по времени только первого порядка. Поэтому для вычисления $G^0_\mu(0, x^i)$ достаточны одни лишь данные Коши $g_{\mu\nu}(0, x^i)$ и $\partial g_{\mu\nu}(0, x^i)/\partial t$ и потому уравнение $G^0_\mu(0, x^i) = 0$ представляет собой условие на данные Коши, необходимое для того, чтобы пространство-время $g_{\mu\nu}(t, x^i)$ имело именно эти данные Коши и удовлетворяло уравнению $G_{\mu\nu} = 0$, которое эквивалентно уравнению $R_{\mu\nu} = 0$.

Формулировка задачи Коши для системы $R_{\mu\nu} = 0$ в части существования решения состоит в следующем.

Пусть $(\mathring{g}_{\mu\nu}(x^i), \mathring{k}_{\mu\nu}(x^i))$ — *данные Коши класса* $(H^s(\Omega), H^{s-1}(\Omega))$, $s \geqslant 3$, *такие, что* $\mathring{G}^0_\mu(x^i) = 0$. *Пусть* Ω_0 — *собственная подобласть*, $\overline{\Omega}_0 \subset \Omega$. *Требуется найти* $\varepsilon > 0$ *и пространство-время* $g_{\mu\nu}(t, x^i)$, $|t| < \varepsilon$, $(x^i) \in \Omega_0 \subset \Omega$, *такие, что*

(I) $g_{\mu\nu}(t, x^i)$ *принадлежит классу* H^s *по совокупности компонент в* $(t, x^i) \in (-\varepsilon, \varepsilon) \times \Omega_0$;

(II) $(g_{\mu\nu}(0, x^i), \partial g_{\mu\nu}(0, x^i)/\partial t) = (\mathring{g}_{\mu\nu}(x^i), \mathring{k}_{\mu\nu}(x^i))$;

(III) $g_{\mu\nu}(t, x^i)$ *обладает нулевой кривизной Риччи.*

Система $R_{\mu\nu} = 0$ является квазилинейной системой десяти дифференциальных уравнений в частных производных, у которых в членах высшего порядка компоненты перемешиваются. В такой постановке для этой системы нет готовых теорем из теории дифференциальных уравнений в частных производных, которые можно было бы применить к сформулированной здесь задаче Коши. Как, однако, было впервые замечено Ланцошем [126] (а для линеаризованных уравнений фактически и самим Эйнштейном [79]), тензор Риччи существенно упрощается в гармонических координатах, т. е. в координатной системе, в которой обращаются в нуль свернутые символы Кристоффеля: $\Gamma^\mu = g^{\alpha\beta}\Gamma^\mu_{\alpha\beta} = 0$. В самом деле, алгебраическим вычислением можно показать, что

$$R_{\mu\nu} = -\frac{1}{2} g^{\alpha\beta} \frac{\partial^2 g_{\mu\nu}}{\partial x^\alpha \partial x^\beta} + \frac{1}{2} g_{\mu\alpha} \frac{\partial \Gamma^\alpha}{\partial x^\nu} + \frac{1}{2} g_{\nu\alpha} \frac{\partial \Gamma^\alpha}{\partial x^\mu} + H_{\mu\nu}$$

и, следовательно, в координатах, для которых $\Gamma^\mu=0$,

$$R_{\mu\nu}=R^{(h)}_{\mu\nu}=-\frac{1}{2}g^{\alpha\beta}\frac{\partial^2 g_{\mu\nu}}{\partial x^\alpha\,\partial x^\beta}+H_{\mu\nu}.$$

Оператор $-(1/2)g^{\alpha\beta}(\partial^2/\partial x^\alpha\partial x^\beta)$ действует одинаково на каждую компоненту системы $g_{\mu\nu}$, и, следовательно, не происходит перемешивания компонент в старших производных. Поэтому нормализованная система $R^{(h)}_{\mu\nu}=0$ значительно проще, чем полная. Действительно, система $R^{(h)}_{\mu\nu}=0$ имеет только простые характеристики, и, следовательно, она является строго гиперболической системой.

Важность использования гармонических координат и системы $R^{(h)}_{\mu\nu}=0$ обусловлена тем, что этого достаточно для решения задачи Коши для уравнения $R^{(h)}_{\mu\nu}=0$. Этот замечательный факт, обнаруженный Шоке-Брюа [32], основан на том наблюдении, что для решений $g_{\mu\nu}$ системы $R^{(h)}_{\mu\nu}=0$ условие $\mathring{\Gamma}^\mu(x^i)=\mathring{g}^{\alpha\beta}(x^i)\mathring{\Gamma}^\mu_{\alpha\beta}(x^i)=0$ распространяется за пределы гиперповерхности $t=0$. Именно это утверждается в следующей лемме.

Лемма 22

Пусть $(\mathring{g}_{\mu\nu}(x^i),\ \mathring{k}_{\mu\nu}(x^i))$ *принадлежат классу Соболева* $(H^s,\ H^{s-1})$ на Ω, $s>(1/2)n+1$, $n=3$, *и предположим, что* $(\mathring{g}_{\mu\nu}(x^i),\ \mathring{k}_{\mu\nu}(x^i))$ *удовлетворяют условиям*

(I) $\mathring{\Gamma}^\mu(x^i)=0$,

(II) $\mathring{G}^0_\mu((x^i)=0$.

Если $g_{\mu\nu}(t,x)$, $|t|<\varepsilon$, $x\in\Omega_0$, Ω_0 — *собственная подобласть*, $\overline{\Omega}_0\subset\Omega$, *является* H^s-*решением системы, то*

$$R^{(h)}_{\mu\nu}=-\frac{1}{2}g^{\alpha\beta}\frac{\partial^2 g_{\mu\nu}}{\partial x^\alpha\,\partial x^\beta}+H_{\mu\nu}=0,$$

$$\left(g_{\mu\nu}(0,x),\ \frac{\partial g_{\mu\nu}(0,x)}{\partial t}\right)=(\mathring{g}_{\mu\nu}(x^i),\ \mathring{k}_{\mu\nu}(x^i)),$$

то $\Gamma^{\mu\nu}(t,x^i)=0$ *при* $|t|<\varepsilon$, $x\in\Omega_0$.

Доказательство. Допустим, что $g_{\mu\nu}(t,x^i)$ удовлетворяет условиям (I), (II) и $R^{(h)}_{\mu\nu}=0$. Тогда непосредственным вычислением убеждаемся, что $\Gamma^\mu(t,x^i)=g^{\alpha\beta}(t,x^i)\Gamma^\mu_{\alpha\beta}(t,x^i)$ удовлетворяет условию $\partial\Gamma^\mu(0,x^i)/\partial t=0$. Из $G^{\mu\nu}{}_{;\nu}=0$ и $R^{(h)}_{\mu\nu}=0$ следует, что Γ^μ удовлетворяет системе линейных уравнений

$$g^{\alpha\beta}\frac{\partial^2\Gamma^\mu}{\partial x^\alpha\,\partial x^\beta}+A^{\beta\mu}_\alpha\left(g_{\mu\nu},\ g^{\mu\nu},\ \frac{\partial g_{\mu\nu}}{\partial x^\lambda}\right)\frac{\partial\Gamma^\alpha}{\partial x^\beta}=0.$$

Эта линейная система имеет вид системы (3), для которой уже до-

казаны теоремы существования и единственности. Таким образом, согласно теореме 15, из того, что $\Gamma^{\mu}(0, x^i)=0$ и $\partial\Gamma^{\mu}(0, x^i)/\partial t=0$, следует, что $\Gamma^{\mu}(t, x^i)=0$.

Согласно этой лемме, H^s-решение системы $R^{(h)}_{\mu\nu}=0$ с фиксированными данными Коши является также решением системы $R_{\mu\nu}=0$ (поскольку $\Gamma^{\mu}(t, x)=0 \Rightarrow R^{(h)}_{\mu\nu}=R_{\mu\nu}$) при условии, что эти данные Коши удовлетворяют требованиям (I) $\mathring{\Gamma}_{\mu}=0$ и (II) $\mathring{G}^0_{\mu}=0$. Как уже отмечалось, условие (II) на данные Коши является необходимым для того, чтобы решение $g_{\mu\nu}(t, x)$ удовлетворяло системе $R_{\mu\nu}=0$. Если же не удовлетворено условие (I), то можно найти набор данных Коши, эволюция которых, согласно уравнению $R^{(h)}_{\mu\nu}=0$, дает пространство-время класса H^s, которое после H^{s+1}-преобразования координат приводит к пространству-времени с исходными данными Коши (см. ниже теорему 26 и работу [84]).

Из теоремы 21 следует, что при $s>2{,}5$ данные Коши класса (H^s, H^{s-1}) приводят к развитию во времени класса H^s и что имеет место коши-устойчивость. Мы можем прийти к этому результату, сводя строго гиперболическую систему $R^{(h)}_{\mu\nu}=0$ к квазилинейной симметричной гиперболической системе первого порядка. Ниже приводится набросок доказательства.

Теорема 23

Пусть Ω — открытая ограниченная область в $\mathbb{R}^3$ с собственной подобластью Ω_0, $\bar{\Omega}_0\subset\Omega$, и пусть $(\mathring{g}_{\mu\nu}(x), \mathring{k}_{\mu\nu}(x))$, $(x^i)\in\Omega$, $0\leqslant\mu$, $\nu\leqslant3$, $1\leqslant i\leqslant3$, принадлежат классу Соболева (H^s, H^{s-1}), $s>2{,}5$. Допустим, что $\mathring{\Gamma}^{\mu}(x^i)=0$ и $\mathring{G}^0_{\mu}(x)=0$. Тогда существуют $\varepsilon>0$ и единственная лоренцева метрика $g_{\mu\nu}(t, x)$, $|t|<\varepsilon$, $(x^i)\subset\Omega_0$, такие, что

(I) $g_{\mu\nu}(t, x^i)$ *принадлежат классу H^s по совокупности* компонент;

(II) $R^{(h)}_{\mu\nu}(t, x^i)=0$;

(III) $(g_{\mu\nu}(0, x^i), \partial g_{\mu\nu}(0, x^i)/\partial t)\equiv(\mathring{g}_{\mu\nu}(x^i), \mathring{k}_{\mu\nu}(x^i))$.

Согласно лемме 22, такая метрика $g_{\mu\nu}(t, x^i)$ удовлетворяет также системе $R_{\mu\nu}(t, x^i)=0$. Более того, $g_{\mu\nu}(t, x^i)$ зависит от $(\mathring{g}_{\mu\nu}(x^i)$, $\mathring{k}_{\mu\nu}(x^i))$ и *непрерывно по топологии (H^s, H^{s-1}). Если $(\mathring{g}_{\mu\nu}(x^i), \mathring{k}_{\mu\nu}(x_i))$ принадлежат классу (C^{∞}, C^{∞}) на Ω, то $g_{\mu\nu}(t, x^i)$ принадлежит классу C^{∞} для всех t, при которых существует это решение.*

Ниже будут рассмотрены решения на всем пространстве $\mathbb{R}^3$ с пространственными асимптотическими условиями.

Выше было показано, как данное утверждение следует непосредственно из теоремы 21 и леммы 22. Чтобы дать иное доказательство, использующее теорему 20, мы преобразуем систему $R^{(h)}_{\mu\nu}=0$ к системе первого порядка, для чего введем 10 новых неизвестных $k_{\mu\nu}=\partial g_{\mu\nu}/\partial t$, 30 новых неизвестных $g_{\mu\nu,i}=\partial g_{\mu\nu}/\partial x^i$ и рассмотрим

квазилинейную систему первого порядка из 50 уравнений:

$$\frac{\partial g_{\mu\nu}}{\partial t} = k_{\mu\nu},$$
$$g^{ij}\left(\frac{\partial g_{\mu\nu,\,i}}{\partial t}\right) = g^{ij}\frac{\partial k_{\mu\nu}}{\partial x^i}, \tag{8}$$
$$-g^{00}\frac{\partial k_{\mu\nu}}{\partial t} = 2g^{0j}\frac{\partial k_{\mu\nu}}{\partial x^j} + g^{ij}\frac{\partial g_{\mu\nu,\,i}}{\partial x^j} - 2H_{\mu\nu}(g_{\mu\nu},\ g_{\mu\nu,\,i},\ k_{\mu\nu}).$$

Мы рассматриваем $H_{\mu\nu}$ как полином по $g_{\mu\nu,i}$ и $k_{\mu\nu}$ и рациональную функцию по $g_{\mu\nu}$ со знаменателем $\det g_{\mu\nu} \neq 0$. В качестве первого шага распространим наши начальные данные на все пространство $\mathbb{R}^3$, скажем, приравнивая их метрике Минковского вне некоторого компактного множества и рассматривая систему (8) на $\mathbb{R}^3$. Отметим, что при такой постановке данные Коши не обязательно должны удовлетворять условию $G^0_\mu = 0$ в переходной области.

Матрица g^{ij} имеет обратную $g_{jk} - (g_{j0}g_{k0}/g_{00})$, т. е. $g^{ij}[g_{jk} - g_{j0}g_{k0}/g_{00})] = \delta^i_k$, и поэтому вторую подсистему из 30 уравнений в (8) можно обратить, что дает

$$\frac{\partial g_{\mu\nu,\,i}}{\partial t} = \frac{\partial k_{\mu\nu}}{\partial x^i}. \tag{9}$$

Для $g_{\mu\nu}$ класса C^2 уравнение (9) означает $g_{\mu\nu,i} = \partial g_{\mu\nu}/\partial x^i$, и, следовательно, система (8) эквивалентна системе $R^{(h)}_{\mu\nu} = 0$.

Пусть

$$u = \begin{pmatrix} g_{\mu\nu} \\ g_{\mu\nu,\,i} \\ k_{\mu\nu} \end{pmatrix}$$

есть 50-компонентный вектор-столбец, где $g_{\mu\nu,i}$ расписывается в виде

$$\begin{pmatrix} g_{00,1} \\ \cdot \\ \cdot \\ \cdot \\ g_{33,1} \\ \cdot \\ \cdot \\ \cdot \\ g_{00,3} \\ \cdot \\ \cdot \\ \cdot \\ g_{33,3} \end{pmatrix}$$

Введем 0^{10} — нулевую матрицу 10×10, I^{10} — единичную матрицу 10×10, а также $A^0(u) = A^0(g_{\mu\nu},\ g_{\mu\nu,i},\ k_{\mu\nu})$ и $A^j(g_{\mu\nu},\ g_{\mu\nu,j},\ k_{\mu\nu})$ —

матрицы 50×50, определяемые выражениями

$$A^0(g_{\mu\nu},\ g_{\mu\nu,i},\ k_{\mu\nu})=\begin{pmatrix} I^{10} & 0^{10} & 0^{10} & 0^{10} & 0^{10} \\ 0^{10} & g^{11}I^{10} & g^{12}I^{10} & g^{13}I^{10} & 0^{10} \\ 0^{10} & g^{12}I^{10} & g^{22}I^{10} & g^{23}I^{10} & 0^{10} \\ 0^{10} & g^{13}I^{10} & g^{23}I^{10} & 0^{10} & 0^{10} \\ 0^{10} & 0^{10} & 0^{10} & 0^{10} & -g^{00}I^{10} \end{pmatrix},$$

$$A^j(g_{\mu\nu},\ g_{\mu\nu,i},\ k_{\mu\nu})=\begin{pmatrix} 0^{10} & 0^{10} & 0^{10} & 0^{10} & 0^{10} \\ 0^{10} & 0^{10} & 0^{10} & 0^{10} & g^{j1}I^{10} \\ 0^{10} & 0^{10} & 0^{10} & 0^{10} & g^{j2}I^{10} \\ 0^{10} & 0^{10} & 0^{10} & 0^{10} & g^{j3}I^{10} \\ 0^{10} & g^{1j}I^{10} & g^{2j}I^{10} & g^{3j}I^{10} & 2g^{j0}I^{10} \end{pmatrix},$$

и пусть $B(g_{\mu\nu},\ g_{\mu\nu,i},\ k_{\mu\nu})$ — 50-компонентный вектор-столбец:

$$B(g_{\mu\nu},\ g_{\mu\nu,i},\ k_{\mu\nu})=\begin{pmatrix} k_{\mu\nu} \\ 0^{30} \\ -2H_{\mu\nu}(g_{\mu\nu},\ g_{\mu\nu,i},\ k_{\mu\nu}) \end{pmatrix}$$

где 0^{30} — 30-компонентный нулевой вектор-столбец.

Отметим, что $A^0(u)$ и $A^j(u)$ симметричны, а $A^0(u)$ положительно определена, если $g_{\mu\nu}$ имеет лоренцеву сигнатуру. Прямое вычисление показывает, что квазилинейная симметричная гиперболическая система

$$A^0(u)\frac{\partial u}{\partial t}=A^j(u)\frac{\partial u}{\partial x^j}+B(u)$$

является как раз системой (8). Из теоремы 20 следует, что для данных Коши

$$\dot u(x^i)=\begin{pmatrix} \dot g_{\mu\nu}(x^i) \\ \dot g_{\mu\nu,i}(x^i) \\ \dot k_{\mu\nu}(x^i) \end{pmatrix}$$

класса Соболева H^{s-1}, $s-1>\frac{1}{2}n+1$ существует $\varepsilon>0$ и решение

$$u(t,\ x^i)=\begin{pmatrix} g_{\mu\nu}(t,\ x^i) \\ g_{\mu\nu,i}(t,\ x^i) \\ k_{\mu\nu}(t,\ x^i) \end{pmatrix}$$

класса H^{s-1}. По лемме Соболева, $u(t,\ x^i)$ принадлежит также классу C^1, и потому из второй группы уравнений в системе (5) следует, что $g_{\mu\nu,i}=\partial g_{\mu\nu}/\partial x^i$. Поскольку $(g_{\mu\nu,i},\ k_{\mu\nu})=(\partial g_{\mu\nu}/\partial x^i,\ \partial g_{\mu\nu}/\partial t)$ принадлежит классу H^{s-1}, $g_{\mu\nu}(t,\ x^i)$ в действительности принадлежит классу H^s. Непрерывная зависимость этих решений от начальных данных следует из общей теории.

Чтобы из результата для $\mathbb{R}^n$ получить результат для области Ω, можно воспользоваться стандартными утверждениями относительно области зависимости (см. [60]).

Поскольку область Ω ограничена, принадлежность $(\overset{\circ}{g}_{\mu\nu}, \overset{\circ}{k}_{\mu\nu})$ классу C^∞ означает, что решение находится на пересечении всех пространств Соболева и, следовательно, также принадлежит классу C^∞; здесь мы снова пользуемся общим результатом о регулярности для симметричных гиперболических систем.

Согласно лемме 22, найденная таким образом метрика $g_{\mu\nu}(t, x^i)$ удовлетворяет уравнению $R_{\mu\nu}=0$.

В то время как из подхода, основанного на системе второго порядка, следует $s>2{,}5$, например $s=3$ [см. теорему 21, (II)], подход, основанный на системе первого порядка, в рассмотренном виде дает $s>3{,}5$, например $s=4$. Результат может быть уточнен, но для этого требуется знание конкретной структуры уравнений и эллиптичность. По этой причине методы, основанные на системе второго порядка, представляются более привлекательными.

В случае асимптотических условий следует проявлять некоторую осторожность. Пространство-время с пространственным поведением $1/r$ не будет принадлежать классу H^s. Зафиксируем фоновое пространство-время $g^b_{\alpha\beta}$ с заданным стремлением к метрике Минковского на бесконечности. Например, коэффициент при $1/r$ может определяться некоторой конкретной массой; $g^b_{\alpha\beta}$ может быть решением типа Шварцшильда со сглаженной сингулярностью при $r=0$.

Введем переменные $u_{\alpha\beta}=g_{\alpha\beta}-g^b_{\alpha\beta}$ и будем решать уравнение для них. В отличие от самой метрики $g_{\alpha\beta}$ переменные $u_{\alpha\beta}$ принадлежат классу H^s.

Наложим на $g_{\alpha\beta}$ следующие условия:

$$g^b_{\alpha\beta}\in C^{s+1}_b(\mathbb{R}^3, \mathbb{R}), \qquad \dot{g}^b_{\alpha\beta}\in H^s(\mathbb{R}^3, \mathbb{R}),$$
$$\frac{\partial g^b_{\alpha\beta}}{\partial x^i}\in H^s(\mathbb{R}^3, \mathbb{R}), \quad 0\leqslant\alpha, \beta\leqslant 3, \quad 1\leqslant i\leqslant 3. \tag{10}$$

В переменных $u_{\alpha\beta}$ уравнения (8) принимают вид (7). Поскольку в коэффициенты при производных второго порядка не входят производные u, требуется лишь, чтобы выполнялось условие $s>(^1/_2)n$.

Введем обозначение $H^s_{g^b_{\alpha\beta}}$ для пространства метрик $g_{\alpha\beta}$, таких, что $g_{\alpha\beta}-g^b_{\alpha\beta}\in H^s$ с соответствующей топологизацией. Тогда теорема 21 сводится к следующей.

Теорема 24

Допустим, что выполнены условия (10). *Тогда для* $s>1{,}5$ *и начальных данных, заданных в некотором шаре вокруг* $(g^b_{\alpha\beta}, \dot{g}^b_{\alpha\beta})$ *в* $H^{s+1}_{g^b_{\alpha\beta}}\times H^s_{g^b_{\alpha\beta}}$, *уравнение* (8) *имеет единственное решение в том же пространстве в интервале времени* $[0, T']$, $T'>0$. *Это решение за-*

висит от начальных данных в этом пространстве непрерывно (т. е. оно корректно, или «коши-устойчиво») и гладко в смысле теоремы 19.

Таким образом, после выделения асимптотических условий начальные данные класса $H^3 \times H^2$ продолжают область пространства-времени класса H^3, причем так, что зависимость от начальных данных непрерывна. Если T' принимает большие значения, то лоренцев характер метрики $g_{\alpha\beta}$ может быть утерян или может появиться сингулярность.

Побочным результатом доказательства данной теоремы является регулярность: если существует решение в $H^{s+1} \times H^s$ на $[0, T']$, то более гладким начальным данным на *том же* интервале $[0, T']$ соответствуют более гладкие решения. Поэтому начальные данные класса C^∞ порождают C^∞-решения.

Интересно было бы установить, является ли пространство, порожденное начальными данными, которые удовлетворяют условию (10), настолько большим, чтобы включать асимптотические бусты. Из анализа доказательств видно, что время существования растет на пространственной бесконечности по крайней мере логарифмически и не похоже, чтобы из этих доказательств следовал утвердительный ответ.

Теперь покажем, что любые два пространства-времени класса H^s, $s>2{,}5$ с нулевой кривизной Риччи и с одними теми же данными Коши связаны координатным H^{s+1}-преобразованием. Ключевая идея состоит в том, чтобы показать, что любое пространство-время класса H^s при введении гармонических координат по-прежнему остается в том же классе. Это в свою очередь базируется на давнем результате Соболева [169], а именно что решения волнового уравнения с коэффициентами класса (H^s, H^{s-1}) сохраняют данные Коши класса (H^{s+1}, H^s); этот результат вытекает из теоремы 15. Мы можем дать альтернативное доказательство этого утверждения, используя известный факт, что любое одно гиперболическое уравнение может быть сведено к системе симметричных гиперболических уравнений (см. [84]). В результате имеем следующее утверждение.

Лемма 25

Пусть $s>2{,}5$ *и* $(\psi_0(x), \dot\psi_0(x))$ *принадлежат классу Соболева* (H^{s+1}, H^s) *на* $\mathbb{R}^3$. *Тогда существует единственная функция* $\psi(t, x)$ *класса* H^{s+1}, *которая удовлетворяет уравнению*

$$g^{\mu\nu}(t, x)\left(\frac{\partial^2\psi}{\partial x^\mu \partial x^\nu}\right) + b^\mu(t, x)\left(\frac{\partial\psi}{\partial x^\mu}\right) + c(t, x)\psi = 0$$

и начальным данным

$$\left(\psi(0, x), \frac{\partial\psi(0, x)}{\partial t}\right) = (\psi_0(x), \dot\psi_0(x)),$$

где $g^{\mu\nu}(t, x)$ — *лоренцева метрика класса* H^s, $b^\mu(t, x)$ — *векторное поле класса* H^{s-1} *и* $c(t, x)$ — *функция класса* H^{s-1}.

Теперь мы можем доказать, что при преобразовании пространства-времени класса H^s к гармоническим координатам оно остается в классе H^s.

Теорема 26

Пусть $g_{\mu\nu}(t, x)$ — *пространство-время класса* H^s; $s>2{,}5$. *Тогда существует координатное* H^{s+1}-*преобразование* $\bar{x}^\lambda(x^\mu)$, *такое, что*

$$\bar{g}_{\mu\nu}(\bar{x}^\lambda)=\frac{\partial x^\alpha}{\partial \bar{x}^\mu}(\bar{x}^\lambda)\frac{\partial x^\beta}{\partial \bar{x}^\nu}(\bar{x}^\lambda)\, g_{\alpha\beta}\,[x^\mu(\bar{x}^\lambda)]$$

является пространством-временем класса H^s, *причем* $\bar{\Gamma}^\mu(\bar{t}, \bar{x})=$ $=\bar{g}^{\alpha\beta}\bar{\Gamma}^\mu_{\alpha\beta}(\bar{t}, \bar{x})=0$.

Доказательство. Чтобы найти $\bar{x}^\lambda(x^\mu)$, рассмотрим волновое уравнение

$$\Box\psi=-g^{\alpha\beta}\left(\frac{\partial^2\psi}{\partial x^\alpha\,\partial x^\beta}\right)+g^{\alpha\beta}\Gamma^\mu_{\alpha\beta}\left(\frac{\partial\psi}{\partial x^\mu}\right)=0$$

и допустим, что $\bar{t}(t, x)$ — единственное решение этого волнового уравнения с данными Коши $\bar{t}(0, x)=0$, $\partial\bar{t}(0, x)/\partial t=1$ и что $\bar{x}^i(t, x)$ — единственное решение того же уравнения с данными Коши

$$\bar{x}^i(0, x)=x^i,\quad \frac{\partial\bar{x}^i}{\partial t}(0, x)=0.$$

Для $g_{\mu\nu}$ класса H^s Γ^μ относится к классу H^{s-1}, и потому $\bar{t}(t, x)$ и $\bar{x}(t, x)$ являются H^{s+1}-функциями; по теореме об обратной функции для H^s-функций [75], $(\bar{t}(t, x), \bar{x}(t, x))$ есть фактически H^{s+1}-диффеоморфизм в окрестности $t=0$.

Поскольку $\Box\bar{x}^\mu(t, x)=0$ есть инвариантное уравнение, в системе координат с чертой имеем

$$\Box\bar{x}^\mu=-\bar{g}^{\alpha\beta}\frac{\partial^2\bar{x}^\mu}{\partial\bar{x}^\alpha\partial\bar{x}^\beta}+\bar{g}^{\alpha\beta}\bar{\Gamma}^\nu_{\alpha\beta}\frac{\partial\bar{x}^\mu}{\partial\bar{x}^\nu}=\bar{g}^{\alpha\beta}\bar{\Gamma}^\mu_{\alpha\beta}=0,$$

и, следовательно, $\bar{x}^\mu$ есть система гармонических координат. ■

Замечание. Эту теорему можно рассматривать как частный случай общей теории гармонических отображений [77].

В качестве простого следствия леммы 25 имеем следующее утверждение о единственности для уравнений Эйнштейна.

Теорема 27

Пусть $g_{\mu\nu}(t, x)$ *и* $\bar{g}_{\mu\nu}(t, x)$ — *два эйнштейновски-плоских* [1]) *пространства-времени класса* H^s *с* $s>2{,}5$, *таких, что* $(g_{\mu\nu}(0, x)$, $\partial g_{\mu\nu}(0, x)/\partial t)=(\bar{g}_{\mu\nu}(0, x)\ \partial\bar{g}_{\mu\nu}(0, x)/\partial t)$; *тогда* $g_{\mu\nu}(t, x)$ *и* $\bar{g}_{\mu\nu}(t, x)$

[1]) То есть с тензором Эйнштейна, равным нулю.— *Прим. перев.*

в некоторой окрестности $t=0$ связаны координатным H^{s+1}-преобразованием.

Доказательство. Согласно лемме 25, существуют координатные H^{s+1}-преобразования $y^\mu(x^\alpha)$ и $\bar{y}^\mu(x^\alpha)$, такие, что преобразованные метрики

$$(\partial x^\alpha/\partial y^\mu)(\partial x^\beta/\partial y^\nu)g_{\alpha\beta} \text{ и } (\partial x^\alpha/\partial \bar{y}^\mu)(\partial x^\beta/\partial \bar{y}^\nu)\bar{g}_{\alpha\beta}$$

удовлетворяют уравнению $R^{(h)}_{\mu\nu}=0$. Поскольку данные Коши для $g_{\mu\nu}$ и $\bar{g}_{\mu\nu}$ совпадают, для преобразованных метрик данные Коши также одинаковы. В силу единственности

$$(\partial x^\alpha/\partial y^\mu)(\partial x^\beta/\partial y^\nu)g_{\alpha\beta} = (\partial x^\alpha/\partial \bar{y}^\mu)(\partial x^\beta/\partial \bar{y}^\nu)\bar{g}_{\alpha\beta}.$$

Поскольку композиция преобразований координат класса H^{s+1} принадлежит тому же классу H^{s+1}, в некоторой окрестности $t=0$ метрика $\bar{g}_{\alpha\beta}$ связана с $g_{\alpha\beta}$ преобразованием координат класса H^{s+1}. ■

Локальные теоремы существования и единственности 23 и 27 могут быть глобализованы в том же духе, как при исследовании максимальных кривых обыкновенных дифференциальных уравнений. Это приводит к следующей теореме, принадлежащей Шоке-Брюа и Героку [50].

Теорема 28

Зафиксируем компактное многообразие M и пусть $(g_0, \pi_0) \in \mathscr{C}_{\mathscr{H}} \cap \cap \mathscr{C}_\delta = \mathscr{C}$ (т. е. являются решением уравнений связи) Тогда существуют пространство-время $(V_4, {}^{(4)}g_0)$ и пространственноподобное вложение i_0: $M \to V_4$, такие, что

(I) $\mathrm{Ein}({}^{(4)}g_0)=0$;

(II) *метрика и сопряженный импульс, индуцированные на* $\Sigma_0 = = i_0(M)$, *совпадают с* (g_0, π_0);

(III) Σ_0 *является поверхностью Коши* [1]);

(IV) $(V_4, {}^{(4)}g_0)$ *является максимальным пространством-временем* [*т. е. не может быть погружено собственным и изометричным образом в другое пространство-время со свойствами* (I), (II) *и* (III)].

Это пространство-время $(V_4, {}^{(4)}g_0)$ *единственно в том смысле, что если имеется другое пространство-время* $(V'_4, {}^{(4)}g'_0)$, *для которого выполняются условия* (I)—(IV), *то существует единственный диффеоморфизм* $F : V_4 \to V'_4$, *такой, что*

(I) $F^* \, {}^{(4)}g'_0 = {}^{(4)}g_0$ (*т. е. F изометричен*);

(II) $F \circ i_0 = i'_0$.

---

[1]) И таким образом, пространство-время $(V_4, {}^{(4)}q)$ глобально-гиперболично ([104], предложение 6.6.3), а потому любая компактная гиперповерхность есть поверхность Коши [24].

Доказательство можно найти в работе [104]. Единственность F доказывается с использованием того факта, что изометрия определена своим действием на систему отсчета в некоторой точке. Линеаризованная версия этого результата понадобится нам в следующем разделе (подробности см. в работе [92]).

Теорема 29

Пусть $(V_4, {}^{(4)}g_0)$ — *вакуумное пространство-время* (*т. е.* $\mathrm{Ein}({}^{(4)}g_0)=0$) *с компактной поверхностью Коши* $\Sigma_0=i_0(M)$ *и с индуцированными метрикой и каноническим импульсом* $(g_0, \pi_0)\in\mathscr{C}_{\mathscr{H}}\cap\mathscr{C}_\delta$. *Пусть* $(h_0, \omega_0)\in S_2\times S_d^2$ *удовлетворяют линеаризованному уравнению связи, т. е.*

$$\mathrm{D}\,\Phi(g_0, \pi_0)\cdot(h_0, \omega_0)=0.$$

Тогда существует ${}^{(4)}h_0\in S_2(V_4)$, *для которого*

$$\mathrm{D}\,\mathrm{Ein}({}^{(4)}g_0)\cdot{}^{(4)}h_0=0$$

и такое, что линеаризованные данные Коши, индуцированные ${}^{(4)}h_0$ *на* Σ_0, *суть* (h_0, ω_0).

Если ${}^{(4)}h_0'$ — *другое такое решение, на* V_4 *существует единственное векторное поле* ${}^{(4)}X$, *такое, что*

$${}^{(4)}h_0'={}^{(4)}h_0+L_{{}^{(4)}X}{}^{(4)}g_0,$$

и ${}^{(4)}X$ *и его производная равны нулю на* Σ_0.

Замечания. а) Линеаризованные данные Коши определяются аналогично определению (g, π), т. е. если ${}^{(4)}g(\rho)$ — кривая лоренцевых метрик, касательных к ${}^{(4)}h$ в точке ${}^{(4)}g_0$, то

$$(h_0, \omega_0)=\left(\frac{\partial g(\rho)}{\partial\rho}\bigg|_{\rho=0}, \frac{\partial\pi(\rho)}{\partial\rho}\bigg|_{\rho=0}\right),$$

где $(g(\rho), \pi(\rho))$ — данные Коши, индуцированные на Σ_0 метрикой ${}^{(4)}g(\rho)$.

б) На гармонические координаты можно смотреть как на техническое средство, с помощью которого проверяется абстрактная теория, изложенная в разд. 3. Однако коль скоро это проделано, корректность имеет место в любой калибровке. Например, можно дать бескоординатную трактовку гиперболических систем (см. [143], с. 247). Более того, при численных расчетах, как показывает работа Смарра и др., разбиения на максимальные слои и слои постоянной средней кривизны могут оказаться более полезными, чем гармонические координаты.

Абстрактная теория, приведенная в разд. 3 (см. теорему 16), относится как к полям, взаимодействующим с гравитацией, так и к чистой гравитации. Однако при этом следует отметить ряд моментов (ср. с работой [104], разд. 7.7):

(I) Взаимодействие полей с гравитацией должно быть минимальным, чтобы не нарушился гиперболический характер уравнений для гравитационного поля.

(II) Тензор энергии-импульса должен быть гладкой функцией (не обязательно полиномиальной) ${}^{(4)}g$, ${}^{(4)}\varphi$.

(III) Для фиксированной ${}^{(4)}g$ (линеаризованные) уравнения материальных полей должны быть корректными. Это необходимо для того, чтобы могла выполняться гипотеза (А1) теоремы 16 [1]).

Остальные условия теоремы 16 носят технический характер, но их нельзя игнорировать (они уточняют условие Б, с. 254 книги Хокинга и Эллиса [104]). Примеры взаимодействующих систем и теорию существования, основанную на прямых методах, читатель найдет в работе [35].

Приведенные выше результаты относительно единственности и глобального развития Коши для вакуумных уравнений стандартным образом переносятся на системы, взаимодействующие с гравитацией.

5. ЛИНЕАРИЗАЦИОННАЯ УСТОЙЧИВОСТЬ ВАКУУМНЫХ УРАВНЕНИЙ ЭЙНШТЕЙНА

Устойчивость линеаризации связана с вопросом о справедливости первого порядка теории возмущений. Суть вопроса состоит в следующем. Допустим, у нас имеются дифференцируемая функция F и точки x_0 и y_0, такие, что $F(x_0)=y_0$. Стандартная процедура нахождения других решений уравнения $F(x)=y_0$ вблизи x_0 состоит в решении линеаризованного дифференциального уравнения $DF(x_0)\cdot h=0$ и в утверждении, что $x=x_0+\rho h$ при малых ρ является приближенным решением уравнения $F(x)=y_0$. Это утверждение может быть уточнено следующим образом: при малых ρ существует кривая точных решений $x(\rho)$, такая, что $F(x(\rho))=y_0$, $x(0)=x_0$ и $x'(0)=h$. Если это утверждение справедливо, мы говорим, что F *линеаризационно устойчива* в точке x_0. Легко указать случай, когда это утверждение неверно. Например, при двух измерениях уравнение $F(x_1, x_2)=x_1^2+x_2^2=0$ не имеет иных решений, кроме (0, 0), хотя линеаризованное уравнение $DF(0, 0)\cdot(h, k)=0$ имеет много решений. Таким образом, обладает или не обладает уравнение линеаризационной устойчивостью вблизи некоторого данного решения, вовсе не праздный вопрос. Интуитивно, линеаризационная устойчивость означает, что теория возмущений первого порядка справедлива вблизи x_0 и не возникает ложных направлений возмущения.

Для общей теории относительности вопрос о линеаризационной устойчивости весьма важен. В литературе часто предполагается, что решения линеаризованных уравнений дают приближения к решениям точных уравнений. Однако Брилл и Дезер [23] показали, что для

[1]) Как отмечается в работе [104], в общих чертах это сводится к утверждению, что «изотропные конусы материальных полей либо совпадают (как, например, в случае системы Эйнштейна — Максвелла) с изотропным конусом пространственно-временной метрики, либо лежат внутри него».

плоского трехмерного тора с нулевой внешней кривизной имеются решения линеаризованных уравнений связи, которые не аппроксимируются кривой точных решений. Они привели доводы, основанные на втором порядке теории возмущений, в пользу того, что при условии $\operatorname{tr}\pi=0$ уравнения связи не имеют иных близлежащих решений, за исключением модификаций, которые по существу тривиальны, хотя у линеаризованных уравнений имеется много нетривиальных решений (полное доказательство см. в работе [88]). Это утверждение аналогично следующей геометрической *теореме изоляции* [89] и доказывается с использованием тех же технических приемов [1]).

Теорема 30

Если M компактно и g_F — плоская метрика на M, то имеется окрестность U_{g_F} метрики g_F в пространстве метрик $\mathcal{M}$, такая, что любая метрика g в окрестности U_{g_F} с $R(g)\geqslant 0$ является плоской.

Доказательство восходит к версии леммы Морса, приспособленной для бесконечномерных пространств; при этом требуется уделить особое внимание координатной инвариантности кривизны.

Результаты по линеаризационной устойчивости получены независимо Шоке-Брюа и Дезером [49] для плоского пространства и Фишером и Марсденом [85, 87, 88] для общего случая пустого пространства-времени с компактной гиперповерхностью. Методы, используемые в этих двух случаях, довольно различны. В работе [154] метод Шоке-Брюа и Дезера обобщен на случай пространства-времени с компактной гиперповерхностью; см. [53]. В работе [61] получены результаты для пространства-времени Робертсона — Уокера, а в работе [2, 3] — для калибровочных полей, взаимодействующих с гравитацией. Результат для плоского пространства состоит в следующем утверждении.

Теорема 31

В окрестности пространства Минковского уравнения Эйнштейна для пустого пространства $\operatorname{Ein}({}^{(4)}g)=0$ *линеаризационно устойчивы.*

Эта теорема связана с необходимостью использования асимптотически плоских пространств и подходящих функциональных пространств с определенными асимптотическими условиями. Здесь мы рассмотрим только компактный случай; некомпактный случай обсуждается в работах [92—94].

Начнем с определения линеаризационной устойчивости уравнений Эйнштейна в пустом пространстве.

---

[1]) Этот результат был недавно глобализован Шёном и Яо как частный случай их решения проблемы массы в общей теории относительности. Например, они доказали, что на трехмерном торе любая метрика с $R(g)\geqslant 0$ является плоской.

Пусть $\mathrm{Ein}({}^{(4)}g_0)=0$. *Инфинитезимальная деформация* метрики ${}^{(4)}g_0$ есть решение ${}^{(4)}h \in S_2(V_4)$ линеаризованных уравнений

$$\mathrm{D\,Ein}({}^{(4)}g_0)\cdot{}^{(4)}h=0.$$

Уравнения Эйнштейна *линеаризационно устойчивы* в окрестности ${}^{(4)}g_0$ (или метрика ${}^{(4)}g_0$ линеаризационно устойчива), если для каждой инфинитезимальной деформации ${}^{(4)}h$ метрики ${}^{(4)}g_0$ существует C^1-кривая ${}^{(4)}g(\rho)$ точных решений уравнений поля в пустом пространстве (на том же самом V_4),

$$\mathrm{Ein}({}^{(4)}g(\rho))=0,$$

таких, что ${}^{(4)}g(0)={}^{(4)}g_0$, $\partial{}^{(4)}g(0)/\partial\rho={}^{(4)}h_0$.

Чтобы быть совершенно строгим, это определение нуждается в небольшом уточнении. А именно, для любого компактного множества $D\subset V_4$ требуется лишь, чтобы ${}^{(4)}g(\rho)$ было определено при $|\rho|<\varepsilon$, где ε может зависеть от D. Это обусловлено тем, что, вообще говоря, ${}^{(4)}g(\rho)$ будет развитием кривой данных Коши $(g(\rho),\ \pi(\rho))$ и поэтому ${}^{(4)}g(\rho)$ при $|\rho|<\varepsilon$ будет равномерно близкой к ${}^{(4)}g_0$ на компактных множествах, но не на всем V_4.

Поскольку мы здесь фиксируем топологию гиперповерхности M, все развития Коши ведут топологически к одному и тому же пространству-времени $V_4\approx\mathbb{R}\times M$, и, следовательно, фиксация V_4 не является серьезным ограничением. Совсем иное дело, конечно, топологические возмущения.

Поскольку используется линеаризованная эйнштейновская динамическая система, линеаризационная устойчивость уравнений Эйнштейна эквивалентна, как мы увидим далее, линеаризационной устойчивости уравнений связи. В действительности, линеаризационная устойчивость корректной гиперболической системы дифференциальных уравнений в частных производных эквивалентна линеаризационной устойчивости любых имеющихся нелинейных связей.

Через линеаризованное отображение $\mathrm{D}\Phi(g,\ \pi)$ можно дать определение необходимого и достаточного условия линеаризационной устойчивости в окрестности $(g_0,\ \pi_0)$ уравнения связи

$$\Phi(g,\ \pi)=0$$

следующим образом: если $(h,\ \omega)\in S_2\times S_d^2$ удовлетворяет линеаризованным уравнениям

$$\mathrm{D}\Phi(g_0,\ \pi_0)\cdot(h,\ \omega)=0,$$

то существует дифференцируемая кривая $(g(\rho),\ \pi(\rho))\in T^*\mathcal{M}$ точных решений уравнений связи

$$\Phi(g(\rho),\ \pi(\rho))=0,$$

такая, что $(g(0), \pi(0))=(g_0, \pi_0)$ и

$$\left(\frac{\partial g(0)}{\partial\rho}, \frac{\partial\pi(0)}{\partial\rho}\right)=(h, \omega).$$

Имеет место следующий основной результат.

Теорема 32

Пусть $\Phi=(\mathscr{H}, \mathscr{Y}) : T^*\mathscr{M}\to C_d^\infty\times\Lambda_d^1$ *определяется как в разд. 2 и, следовательно,* $\mathscr{C}_{\mathscr{H}}\cap\mathscr{C}_\delta=\Phi^{-1}(0)$. *Пусть также* $(g_0, \pi_0)\in\mathscr{C}_{\mathscr{H}}\cap\mathscr{C}_\delta$. *Следующие условия являются эквивалентными:*

(I) *Уравнения связи*

$$\Phi(g, \pi)=0$$

линеаризационно устойчивы в окрестности (g_0, π_0).

(II) *Отображение* $D\Phi(g_0, \pi_0) : S_2\times S_d^2\to C_d^\infty\times\Lambda_d^1$ *сюръективно.*

(III) *Отображение* $[D\Phi(g_0, \pi_0)]^* : C^\infty\times\mathscr{X}\to S_d^2\times S_2$ *инъективно.*

Замечание. В разд. 2 мы привели несколько достаточных условий, при которых выполняется (II),— это условия $C_{\mathscr{H}}$, C_δ и C_{tr}.

Доказательство теоремы 32. В разд. 2 была показана эллиптичность $[D\Phi(g_0, \pi_0)]^*$. Поэтому эквивалентность условий (II) и (III) является непосредственным следствием альтернативы Фредгольма.

Из условия (II) следует (I). Ядро отображения $D\Phi(g_0, \pi_0)$ расщепляется по альтернативе Фредгольма. Поэтому из теоремы о неявной функции следует, что вблизи (g_0, π_0) $\Phi^{-1}(0)$ является гладким многообразием. Теперь мы должны использовать пространства Соболева и перейти к C^∞, накладывая требования регулярности, как это делается в работе [89]. Поскольку любой вектор, касательный к гладкому многообразию, касателен и к кривой на этом многообразии, отсюда вытекает (I).

Из условия (I) следует (III). Это утверждение не столь элементарно, и его доказательство мы кратко приведем. Предположим, что имеет место (I) и что $[D\Phi(g_0, \pi_0)]^*\cdot(N, X)=0$, но $(N, X)\neq 0$. Мы придем к противоречию, показав, что имеется необходимое условие второго порядка на деформацию (h, ω) первого порядка, которое должно быть удовлетворено для того, чтобы деформация была касательной к кривой точных решений уравнений связи. Итак, пусть (h, ω) — решение линеаризованных уравнений связи и $(g(\rho), \pi(\rho))$ — кривая точных решений уравнения

$$\Phi(g(\rho), \pi(\rho))=0, \qquad (11)$$

проходящая через (g_0, π_0) и касательная к (h, ω). Дифференцируя дважды уравнение (11) и полагая $\rho=0$, получаем

$$D\Phi(g_0, \pi_0)\cdot(g''(0), \pi''(0))+D^2\Phi(g_0, \pi_0)\cdot((h, \omega), h, \omega)=0, \qquad (12)$$

где

$$g''(0)=\frac{\partial^2 g(0)}{\partial\rho^2},\quad \pi''(0)=\frac{\partial^2\pi(0)}{\partial\rho^2}.$$

Свернув (12) с (N, X) и интегрируя по M, для первого слагаемого в (12) получим

$$\int\langle(N, X),\ D\Phi(g_0, \pi_0)\cdot(g''(0), \pi''(0))\rangle=$$
$$=\int\langle[D\Phi(g_0, \pi_0)]^*\cdot(N, X),\ (g''(0), \pi''(0))\rangle=0,$$

поскольку $(N, X)\in\ker[D(g_0, \pi_0)]^*$.

Таким образом, первый член в (12) выпадает и остается необходимое условие

$$\int\langle(N, X),\ D^2\Phi(g_0, \pi_0)\cdot((h, \omega), (h, \omega))\rangle=0, \tag{13}$$

которое должно выполняться для всех $(h, \omega)\in\ker D\Phi(g_0, \pi_0)$. Чтобы показать нетривиальность этого условия (см. [95]), можно провести рассуждение, аналогичное тому, которое имеется в работе [19].

Процедура нахождения условия второго порядка, при котором линеаризационная устойчивость отсутствует, вполне обычна. Прочие приложения см. в работах [88, 89]. ■

Из линеаризационной устойчивости уравнений связи можно вывести линеаризационную устойчивость пространства-времени, и наоборот, следующим образом.

Теорема 33

Пусть $(V_4, {}^{(4)}g_0)$ — вакуумное пространство-время, которое является максимальным развитием данных Коши (g_0, π_0) на компактной гиперповерхности $\Sigma_0=i_0(M)$.

Тогда уравнения Эйнштейна на V_4

$$\mathrm{Ein}({}^{(4)}g)=0$$

линеаризационно устойчивы в окрестности ${}^{(4)}g_0$, если и только если уравнения связи

$$\Phi(g, \pi)=0$$

линеаризационно устойчивы в окрестности (g_0, π_0).

В частности, если (g_0, π_0) удовлетворяют условиям $C_{\mathscr{H}}$, C_δ и C_{tr}, то уравнения Эйнштейна линеаризационно устойчивы.

Доказательство. Предположим сначала, что уравнения связи линеаризационно устойчивы. Пусть ${}^{(4)}h_0$ есть решение линеаризованного уравнения в «точке» ${}^{(4)}g_0$ и пусть (h_0, ω_0) — индуцированная деформация (g, π) на Σ_0. При этом (h_0, ω_0) удовлетворяют линеаризованным уравнениям связи. По предположению существует кривая $(g(\rho), \pi(\rho))\in\mathscr{C}_{\mathscr{H}}\cap\mathscr{C}_\delta$, касательная к (h_0, ω_0) в (g_0, π_0).

Согласно теории существования для задачи Коши, имеется кри-

вая ${}^{(4)}g(\rho)$ максимальных решений уравнений $\mathrm{Ein}({}^{(4)}g(\rho))=0$ на $V_4\approx\mathbb{R}\times M$ с данными Коши $(g(\rho),\ \pi(\rho))$. По теоремам 19 и 24, при данном выборе длительности и сдвига ${}^{(4)}g(\rho)$ будет гладкой функцией ρ в смысле теоремы 19 или в обычном смысле принадлежности к классу C^∞. Как и раньше, для любого компактного множества $D\subset\subset V_4$ и $\varepsilon>0$ найдется $\delta>0$, такое, что ${}^{(4)}g(\rho)$ находится в ε-окрестности ${}^{(4)}g_0$ (по любой стандартной топологии) на D.

Пользуясь результатами по единственности для линеаризованной и полной систем Эйнштейна, можно преобразовать кривую ${}^{(4)}g(\rho)$ диффеоморфизмами так, что ${}^{(4)}h_0$ будет касаться ее при $\rho=0$. Подробности см. в [92]. ■

Монкри [145] доказал, что для $(g,\ \pi)\in\mathscr{C}_{\mathscr{H}}\cap\mathscr{C}_\delta$ отображение $[D\Phi(g,\ \pi)]^*$ инъективно, если и только если пространство-время ${}^{(4)}g$, порожденное $(g,\ \pi)$, не имеет никаких (нетривиальных) векторных полей Киллинга ${}^{(4)}Y$ (т. е. из $L_{{}^{(4)}Y}\,{}^{(4)}g=0$ следует, что ${}^{(4)}Y=0$); вместе с теоремами 32 и 33 результат Монкри дает тогда необходимое и достаточное условие линеаризационной устойчивости пространства-времени с компактными пространственноподобными гиперповерхностями Коши.

Результат Монкри все же не обеспечивает необходимых и достаточных условий инъективности $[D\Phi(g,\ \pi)]^*$ через $(g,\ \pi)$ (условия $C_{\mathscr{H}}$, C_δ и C_{tr} достаточны, но не необходимы), но зато обходится без условия $\mathrm{tr}\,\pi'=\mathrm{const}$, которое, таким образом, становится менее важным.

Теорема 34 [145]

Пусть ${}^{(4)}g$ *— решение уравнений Эйнштейна в пустом пространстве* $\mathrm{Ein}({}^{(4)}g)=0$. *Пусть* $\Sigma_0=i_0(M)$ *— компактная гиперповерхность Коши с индуцированной метрикой* g_0 *и каноническим импульсом* π_0. *Тогда* $\ker\,[D\Phi(g_0,\ \pi_0)]^*$ *(конечномерное векторное пространство) изоморфно пространству векторных полей Киллинга пространства-времени* ${}^{(4)}g$. *В действительности*

$$(Y_\perp,\ Y_\parallel)\in\ker[D\Phi(g_0,\ \pi_0)]^*,$$

если и только если в ${}^{(4)}g$ *существует векторное поле Киллинга* ${}^{(4)}Y$, *нормальными и касательными* к Σ_0 *составляющими которого являются* $Y_\perp$ *и* $Y_\parallel$.

Доказательства, альтернативные доказательству Монкри, см. в работах [56—58, 92].

В качестве важного следствия этого результата мы замечаем, что *условие* $\ker\,[D\Phi(g_0,\ \pi_0)]^*=\{0\}$ *не зависит от гиперповерхности* Σ_0 (поскольку оно эквивалентно отсутствию векторных полей Киллинга, что не зависит от гиперповерхности). Это условие, очевидно, не меняется также при переходе к изометрическому пространству-времени.

В результате приходим к следующей основной теореме о линеаризационной устойчивости.

Теорема 35

Пусть ${}^{(4)}g_0$ *— решение вакуумных уравнений Эйнштейна* $\mathrm{Ein}({}^{(4)}g_0)=0$. *Предположим, что пространство-время* $(V_4, {}^{(4)}g_0)$ *имеет компактную поверхность Коши* Σ_0.

Тогда уравнения Эйнштейна на V_4

$$\mathrm{Ein}({}^{(4)}g)=0$$

линеаризационно устойчивы в «точке» ${}^{(4)}g_0$, *если и только если* ${}^{(4)}g_0$ *не допускает ни одного векторного поля Киллинга.*

Мы завершим этот раздел кратким рассмотрением случая, когда ${}^{(4)}g_0$ не обладает линеаризационной устойчивостью. Цель состоит в том, чтобы найти те необходимые и достаточные условия на решение ${}^{(4)}h$ линеаризованных уравнений, при которых ${}^{(4)}h$ касается кривой точных решений, проходящей через ${}^{(4)}g_0$. Необходимые условия будут выведены; относительно достаточности см. [95].

В теореме 32 мы показали, что если ${}^{(4)}h$ касается кривой точных решений и $(N, X)\in \ker[\mathrm{D}\Phi(g_0, \pi_0)]^*$, то

$$\int_{\Sigma_0} \langle (N, X), \mathrm{D}^2\Phi(g_0, \pi_0)\cdot((h, \omega), (h, \omega))\rangle = 0.$$

Следуя [148], мы можем переписать это условие второго порядка через метрику пространства-времени точно так, как уже было переписано условие $\ker \mathrm{D}\Phi(g_0, \pi_0)=\{0\}$ (см. альтернативные доказательства в работах [92, 95]).

Теорема 36 [148]

Пусть $\mathrm{Ein}({}^{(4)}g_0)=0$ *и* ${}^{(4)}h\in S_2(V_4)$ *удовлетворяет линеаризованным уравнениям*

$$\mathrm{D}\,\mathrm{Ein}({}^{(4)}g_0)\cdot{}^{(4)}h=0.$$

Пусть ${}^{(4)}Y$ *— векторное поле Киллинга метрики* ${}^{(4)}g_0$ *(и, следовательно,* ${}^{(4)}g_0$ *линеаризационно неустойчива). Пусть* Σ_0 *— компактная гиперповерхность Коши и пусть* $(Y_\perp, Y_\parallel)$ *— нормальная и касательная компоненты* ${}^{(4)}Y$ *на* Σ_0. *Тогда необходимым условием второго порядка для того, чтобы* ${}^{(4)}h$ *было касательно к кривой точных решений, является равенство*

$$\int_{\Sigma_0} \langle \mathrm{D}^2\,\mathrm{Ein}({}^{(4)}g_0)\cdot({}^{(4)}h, {}^{(4)}h), ({}^{(4)}Y_{\Sigma_0}, {}^{(4)}Z_{\Sigma_0})\rangle\, d\mu(g_0) =$$

$$= \int_{\Sigma_0} \langle (Y_\perp, Y_\parallel), \mathrm{D}^2\Phi(g_0, \pi_0)\cdot((h, \omega), (h, \omega))\rangle = 0. \qquad (14)$$

Если $\mathrm{Ein}({}^{(4)}g_0)=0=\mathrm{D}\,\mathrm{Ein}({}^{(4)}g_0)\cdot{}^{(4)}h_0$, то $\mathrm{D}^2\mathrm{Ein}({}^{(4)}g_0)\cdot({}^{(4)}h, {}^{(4)}h)$ имеет дивергенцию, равную нулю [171]. Таким образом, если ${}^{(4)}Y$ — векторное поле Киллинга, то дивергенция векторного поля

$${}^{(4)}W = {}^{(4)}Y\cdot[\mathrm{D}^2\,\mathrm{Ein}({}^{(4)}g_0)\cdot({}^{(4)}h, {}^{(4)}h)]$$

также равна нулю. Следовательно, необходимое условие второго порядка

$$\int_{\Sigma_0} \langle {}^{(4)}W, {}^{(4)}Z_{\Sigma_0} \rangle d\mu(g_0) = 0$$

на деформацию первого порядка не зависит от гиперповерхности Коши, на которой оно накладывается. Тогда интеграл от ${}^{(4)}W$ по гиперповерхности Коши представляет собой сохраняющуюся величину для гравитационного поля, построенную из решения ${}^{(4)}h$ линеаризованного уравнения и векторного поля Киллинга ${}^{(4)}Y$. Интересным и важным свойством этой сохраняющейся величины Тауба является, как это видно из теоремы 36, тот факт, что решение первого порядка ${}^{(4)}h$, из которого построено ${}^{(4)}W$, не является касательным к какой-либо кривой точных решений, если эта величина не равна нулю. Таким образом, для пространства-времени, которое не является линеаризационно устойчивым, сохраняющая величина Тауба играет решающую роль для ответа на вопрос о допустимости возмущения ${}^{(4)}h$ (т. е. о том, является ли оно касательным к какой-либо кривой точных решений).

6. ПРОСТРАНСТВО ГРАВИТАЦИОННЫХ СТЕПЕНЕЙ СВОБОДЫ

Теперь мы дадим обзор некоторых результатов симплектической геометрии, которая служит основой для единообразного описания различных расщеплений, возможных в общей теории относительности [4]. Эти результаты основаны на общей редукции фазовых пространств, для которых имеется инвариантная относительно действия некоторой группы гамильтонова система [142]. Последующее применение этих результатов ведет к построению симплектического пространства гравитационных степеней свободы [93] [1]). Основополагающими по содержанию данного раздела являются работы [1, 30, 138].

Пусть P — некоторое многообразие и Ω — симплектическая форма на P, т. е. Ω есть замкнутая (слабо) невырожденная 2-форма. В случае общей теории относительности P есть $T^*\mathcal{M}$ и Ω — каноническая симплектическая форма J^{-1}, уже введенные в разд. 1.

Пусть G — топологическая группа, действующая канонически на P, т. е. для любого $g \in G$, $\Phi_g: p \mapsto g \cdot p$, действие g на P сохраняет Ω. Допустим, что для этого действия существует *импульс* Ψ. Это означает, что Ψ есть отображение из P в $\mathfrak{g}^*$ ($\mathfrak{g}^*$ дуально к алгебре Ли $\mathfrak{g} = T_eG$ группы G), такое, что

$$\Omega(\xi_P(p), v_p) = \langle d\Psi(p) \cdot v_p, \xi \rangle$$

---

[1]) Следует отметить, что в случае компактных поверхностей Коши пространство гравитационных степеней свободы отфакторизовывает все динамические степени свободы. Для некоторых целей это может оказаться нежелательным и может понадобиться менее жесткое отождествление (см. [91, 177, 178]).

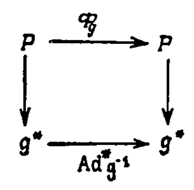
для всех $\xi \in \mathfrak{g}$, где ξ_P — соответствующий инфинитезимальный генератор (форма Киллинга) на P и $v_p \in T_pP$. При ином способе введения Ψ от отображения $p \mapsto \langle \Psi(p), \xi \rangle$ требуется, чтобы для всякого ξ оно было энергетической функцией гамильтонова векторного поля ξ_P. Такое понятие импульса представляет собой существенную геометризацию различных теорем сохранения классической механики и теории поля, включая теорему Нётер.

Легко доказать следующее утверждение: если H — гамильтонова функция на P с соответствующим гамильтоновым векторным полем X_H, т. е. $dH(p)\cdot v = \Omega_p(X_H(p), v)$, или, что то же самое, $i_{X_H}\Omega = \mathrm{d}H$, и если H инвариантна относительно G, то Ψ есть интеграл движения для X_H, т. е. если F_t есть поток X_H, то $\Psi \circ F_t = \Psi$.

В качестве примера рассмотрим группу G, действующую на конфигурационном пространстве Q. Лифт действия этой группы в фазовое пространство T^*Q является каноническим преобразованием. Импульс в этом случае определяется равенством

$$\langle \Psi(\alpha_q), \xi \rangle = \langle \xi_Q(q), \alpha_q \rangle,$$

где α_q принадлежит T^*Q. Если G есть множество трансляций или вращений, то Ψ является импульсом или моментом импульса соответственно. Как и следовало ожидать, Ψ есть вектор, и трансформационным свойством, требующимся от этого вектора, является эквивариантность импульса или момента относительно косопряженного действия G на $\mathfrak{g}$, т. е. диаграмма



должна быть коммутативной. Мы будем рассматривать лишь эквивариантные импульсы и моменты.

Имеется несколько классических теорем относительно редукции фазовых пространств. В небесной механике есть теорема Якоби об исключении узлов орбит, утверждающая, что во вращательно-инвариантной системе из общего числа переменных можно исключить четыре переменные и при этом в новых координатах система останется гамильтоновой. Другая классическая теорема гамильтоновой механики утверждает, что существование k первых интегралов в инволюции позволяет сократить $2k$ переменных в фазовом пространстве. Обе эти теоремы следуют из теоремы Марсдена и Вайнштейна [142] о редукции фазового пространства.

Для построения этого редуцированного пространства положим $\mu \in \mathfrak{g}^*$ и

$$G_\mu = \{g \in G \mid \mathrm{Ad}^*_{g^{-1}}\mu = \mu\}.$$

Рассмотрим $\Psi^{-1}(\mu)=\{p|\Psi(p)=\mu\}$. Условие эквивариантности означает, что G сохраняет $\Psi^{-1}(\mu)$, и потому мы можем рассматривать $P_\mu=\Psi^{-1}(\mu)/G_\mu$. В случае когда $\Psi^{-1}(\mu)$ представляет собой многообразие (т. е. μ является регулярной точкой) и действие G на это многообразие является свободным и собственным, имеет место следующая теорема.

Теорема 37

P_μ наследует от P естественную симплектическую структуру, и гамильтонова система на P, инвариантная относительно канонического действия G, естественным образом проецируется на некоторую гамильтонову систему на P_μ.

В теореме Якоби об исключении узлов группы G есть $SO(3)$, так что $\mathfrak{g}$ есть $\mathbb{R}^3$ и косопряженное действие совпадает с обычным. Отсюда подгруппа изотропии G_μ точки μ в $\mathbb{R}^3$ есть S^1. Если размерность P равна n, то $\Psi^{-1}(\mu)$ есть множество решений для трех уравнений, и, следовательно, размерность $\Psi^{-1}(\mu)/G_\mu$ равна $n-3-1=$ $=n-4$. Для k первых интегралов в инволюции группы G есть k-мерная абелева группа, поэтому косопряженное действие тривиально и $G_\mu=G$. Отсюда, размерность $\Psi^{-1}(\mu)/G$ равна $n-2k$. Другой известной теоремой, вытекающей из теоремы 37, является теорема Костанта — Кириллова, утверждающая, что орбита точки $\mu\in\mathfrak{g}^*$ при сопряженном действии есть симплектическое многообразие.

Теперь мы покажем, как получить общую теорему расщепления для симплектических многообразий, одна часть которых касательна к редуцированному пространству P_μ [4]. Она включает как частный случай теоремы расщепления для симметричных тензоров.

Теорема расщепления для симплектического многообразия P требует положительно-определенной, но, возможно, только слабоневырожденной метрики или иной подобной структуры, задающей дуализацию. Если эти требования выполнены, то могут быть введены ортогональные дополнения. Допустим, нам известно, скажем из теоремы Фредгольма, что

$$T_pP=\operatorname{im}(T_p\Psi)^*\oplus\ker T_p\Psi$$

(здесь $(T_p\Psi)^*$ — обычное L_2-сопряжение). Ясно, что при конечном числе измерений это выполняется автоматически. Введем

$$\alpha_p:\mathfrak{g}_\mu\to T_pP;\ \xi\mapsto\xi_P(p),$$

где $\mathfrak{g}_\mu$ — алгебра Ли группы G_μ. Допустим также, что имеется расщепление

$$T_pP=\operatorname{im}\alpha_p\oplus\ker\alpha_p^*.$$

Для этих двух расщеплений существует общее условие совместности, а именно $\operatorname{im}\alpha_p\subset\ker T_p\Psi$, которое прямо следует из эквивари

антности. В самом деле,

$$\operatorname{im}\alpha_p = T_p(G\cdot p)\cap\ker T_p\Psi.$$

Это условие совместности приводит к более тонкому расщеплению

$$T_pP = \operatorname{im}(T_p\Psi)^*\oplus\operatorname{im}\alpha_p\oplus(\ker T_p\Psi\cap\ker\alpha_p^*), \qquad (15)$$

т. е.

$$T_pP = \operatorname{im}(T_p\Psi)^*\oplus T_p(G_\mu\cdot p)\oplus\ker T_p\Psi/[T_p(G_\mu\cdot p)].$$

Заметим, что третье слагаемое есть касательное пространство к P_μ.

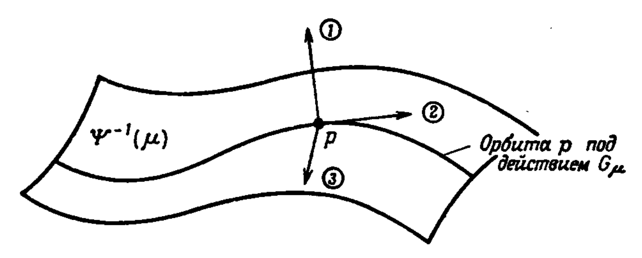


Рис. 2. Геометрия общего симплектического разложения.

Геометрическая картина показана на рис. 2. Для пояснения этого рисунка перечислим слагаемые в разложении (15) в виде

$$T_pP = ①\oplus②\oplus③,$$

где

① принадлежит $\operatorname{im}(T_p\Psi)^*$, ортогональному дополнению касательного пространства к уровню $\Psi^{-1}(\mu)$;

② принадлежит $\operatorname{im}\alpha_p$, касательному пространству орбиты точки p под действием G_μ;

③ лежит в $\ker T_p\Psi\cap\ker\alpha_p^*$ и является частью разложения, касательного к редуцированному симплектическому многообразию.

② и ③ вместе составляют $\ker T_p\Psi$, касательное пространство к $\Psi^{-1}(\mu)$.

Основное расщепление в [146] можно рассматривать как частный случай этого результата. Мы выберем $P=T^*\mathcal{M}$, а «группой» будет тогда $G=C^\infty_{\text{простр}}(M;\ V_4,\ {}^{(4)}g)$, множество пространственноподобных вложений M в гиперповерхности Коши в пространство-время с нулевым тензором Эйнштейна $(V_4,\ {}^{(4)}g)$, которое является максимальным развитием с некоторой гиперповерхности Коши $\Sigma\subset V_4$.

Хотя G группой не является, но имеет достаточно сходства с группой, чтобы был применим наш анализ [1]). G «действует» на (g,π) следующим образом (рис. 3). Пусть $(V_4,\ {}^{(4)}g,\ i_0)$, $\operatorname{Ein}({}^{(4)}g)=0$ есть максимальное развитие данных Коши, которым обладают $(g_0,\ \pi_0)$

[1]) Нужно лишь использовать более общую процедуру редукции, описанную в работах [1, 173].

как данные Коши на некоторой вложенной гиперповерхности $\Sigma_0 = i_0(M)$, $i_0 : M \to V_4$ (i_0 уподобляется началу координат для $C^\infty_{\text{простр}}$). Тогда $i \in C^\infty_{\text{простр}}(M;\ V_4,\ {}^{(4)}g)$ отображает (g_0, π_0) на (g, π), индуцированные на гиперповерхности $\Sigma = i(M)$. Множество всех таких (g, π) определяет орбиту (g_0, π_0) в $\mathscr{C}_{\mathscr{H}} \cap \mathscr{C}_\delta$. Эти орбиты не имеют общих точек и потому определяют отношение эквивалентности $(\sim)$ в $\mathscr{C}_{\mathscr{H}} \cap \mathscr{C}_\delta$.

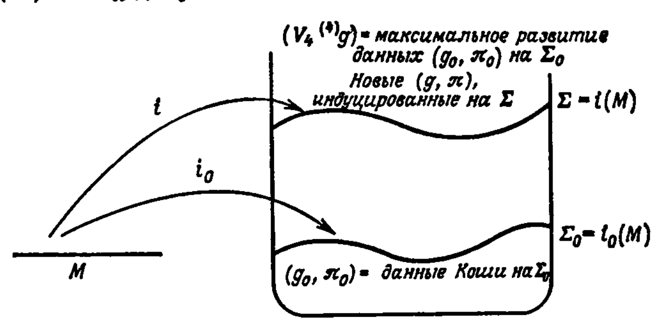


Рис. 3. Схема «действия» пространства вложений на пространстве данных Коши.

Хотя это не есть действие (поскольку $C^\infty_{\text{простр}}$ не является группой), но все же здесь есть хорошо определенные орбиты и применим изложенный выше симплектический анализ [93]. С помощью сопряженной формы эволюционной системы Эйнштейна вычисляется импульс этого «действия» на касательный вектор ${}^{(4)}X_\Sigma \in T_{i_0}C^\infty_{\text{простр}}(M;\ V_4,\ {}^{(4)}g)$ с функциями длительности N и сдвига X:

$$\Psi_{(g,\pi)}({}^{(4)}X) = \int N\mathscr{H}(g, \pi) + X \cdot \mathscr{Y}(g, \pi).$$

Здесь ${}^{(4)}X_\Sigma$ или (N, X) можно рассматривать как элементы «алгебры Ли» «группы» $C^\infty_{\text{простр}}$.

Поскольку $\Psi^{-1}(0)$ совпадает в точности с множеством связей $\mathscr{C}_{\mathscr{H}} \cap \mathscr{C}_\delta$, мы выберем $\mu = 0$, и, следовательно, $G_\mu = G$. Из уравнений движения находим, что

$$\alpha_{(g,\pi)} : \mathfrak{g} \to T_{(g,\pi)}(T^*\mathscr{M})$$

определяется через

$$(N,\ X) \mapsto J \circ [D\Phi(g,\ \pi)]^* \cdot \begin{pmatrix} N \\ X \end{pmatrix},$$

и поэтому симплектическое разложение (15) принимает вид

$$\begin{gathered} T_{(g,\pi)}(T^*\mathscr{M}) = \{\operatorname{im}[D\Phi(g,\ \pi)]^*\}^* \oplus \\ \oplus \operatorname{im}\{J \circ [D(g,\ \pi)]^*\} \oplus \\ \oplus \ker D\Phi(g,\ \pi) \cap [\ker D\Phi(g,\ \pi) \circ J]^*, \end{gathered}$$

т. е. представляет собой расщепление Монкри. Элементы первого слагаемого инфинитезимально деформируют (g, π) в данные Коши,

которые не удовлетворяют уравнениям связи. Элементы второго слагаемого инфинитезимально деформируют (g, π) в данные Коши, которые порождают некоторое изометрическое пространство-время, а элементы третьего слагаемого инфинитезимально деформируют (g, π) в направлении новых данных Коши, порождающих неизометрическое решение уравнения поля в пустом пространстве (рис. 4, ср. с рис. 2). Это третье слагаемое представляет собой касательное пространство к редуцированному пространству $P_\mu \approx \mathscr{C}_{\mathscr{H}} \cap \mathscr{C}_\delta / \sim$.

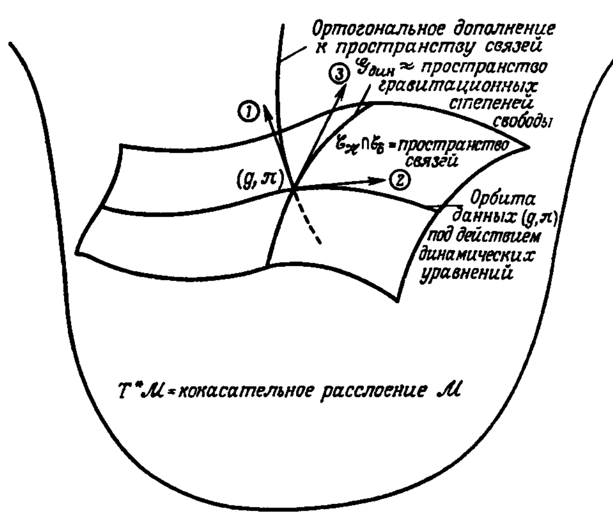


Рис. 4. Симплектическое разложение, примененное к уравнениям Эйнштейна для построения пространства гравитационных степеней свободы.

Существует единственный изоморфизм между этим фактор-пространством (по отношению эквивалентности, введенному выше) и пространством гравитационных степеней свободы

$$\mathscr{G}(V_4) = \mathscr{E}(V_4)/\mathscr{D}(V_4),$$

т. е. множеством максимальных решений вакуумных уравнений Эйнштейна

$$\mathscr{E}(V_4) = \{{}^{(4)}g \mid \mathrm{Ein}({}^{(4)}g) = 0 \text{ и такие, что } (V_4, {}^{(4)}g) \text{ есть максимальное развитие данных Коши на некоторой гиперповерхности Коши}\}$$

по модулю группа пространственно-временных диффеоморфизмов $\mathscr{D}(V_4)$. Оно является пространством изометрических классов решений уравнений Эйнштейна в пустом пространстве или пространством гравитационных степеней свободы, поскольку координатная калибровочная группа отфакторизована.

Представление $\mathscr{G}(V_4)$, описанное здесь, мы называем *динамическим представлением*, так как оно использует каноническую

формулировку для определения $P_\mu \approx \mathscr{C}_{\mathscr{K}} \cap \mathscr{C}_\delta/\sim$. Конформное представление $\mathscr{G}(V_4)$ см. в работах [53, 91, 176].

Как мы уже подчеркивали, в случае компактных гиперповерхностей отождествляются все (g, π), которые возникают при разбиениях в максимальном пространстве-времени с нулевым тензором Эйнштейна. В некомпактном случае, как было показано в работах [54, 161], этого *не* происходит.

Относительно разложения $T_{(g,\pi)}(T^*\mathscr{M})$ следует сделать несколько дополнительных замечаний.

Положим $\mathfrak{g}_{(g,\pi)} = \ker D\Phi(g,\pi) \cap [\ker D\Phi(g,\pi)\circ J]^*$ (третье слагаемое в вышеприведенном разложении Монкри). Это слагаемое обобщает классическое поперечно-бесследовое разложение (TT) в работах Дезера и Брилла [62, 22]. В самом деле, для $\pi = 0$ и $R(g) = 0$ разложение Монкри сводится к двум копиям расщепления Берже — Эбена [16]. Если еще и $\mathrm{Ric}(g) = 0$ (т. е. g — плоская метрика), то воспроизводится первоначальное расщепление Брилла—Дезера.

Теперь допустим, что $(h, \omega) \in \mathscr{G}_{(g,\pi)}$. Тогда (h, ω) удовлетворяет следующим уравнениям:

$$D\Phi(g, \pi)\cdot(h, \omega) = 0, \tag{16}$$

$$[D\Phi(g, \pi)\circ J]\cdot(h, \omega)^* = D\Phi(g, \pi)\cdot((\omega')^\flat, -h^\# d\mu(g)) = 0. \tag{17}$$

Те же уравнения, записанные через функции связи $\mathscr{H}$ и $\mathscr{J}$, имеют вид

$$D\mathscr{H}(g, \pi)\cdot(h, \omega) = 0,$$
$$D\mathscr{H}(g, \pi)\cdot((\omega')^\flat, -h^\# d\mu(g)) = 0,$$
$$D\mathscr{J}(g, \pi)\cdot(h, \omega) = 0,$$
$$D\mathscr{J}(g, \pi)\cdot((\omega')^\flat, -h^\# d\mu(g)) = 0.$$

Эти уравнения, представляющие собой восемь условий на двенадцать функций трех переменных, формально оставляют в качестве параметров пространства $\mathscr{G}_{(g,\pi)}$ четыре функции трех переменных. Формально $\mathscr{G}_{(g,\pi)}$ является касательным пространством к пространству гравитационных степеней свободы, параметризованным четырьмя функциями трех переменных.

Более того, в слагаемом $\mathscr{G}_{(g,\pi)}$ имеется определенная «симплектическая симметрия», которая отражена в равенствах (16) и (17): если $h, \omega \in \mathscr{G}_{(g,\pi)}$, то $J\circ(h, \omega)^*$ также принадлежит $\mathscr{G}_{(g,\pi)}$. Мы будем называть эту симметрию J-инвариантностью пространства $\mathscr{G}_{(g,\pi)}$.

Предложение 38

(Слабая) симплектическая форма Ω на $S_2 \times S_d^2$ естественным образом индуцирует слабую симплектическую форму Ω' на любом J-инвариантном подпространстве в $S_2 \times S_d^2$. В частности, $\mathscr{G}_{(g,\pi)}$ является (слабо) симплектическим линейным пространством.

Доказательство. Симплектическая форма Ω на $S_2 \times S_d^2$, определяемая как

$$\Omega((h_1, \omega_1), (h_2, \omega_2)) = \int_M \langle J^{-1}(h_1, \omega_1), (h_2, \omega_2)\rangle,$$

задает той же формулой и антисимметричную билинейную форму Ω' на $\mathscr{G}_{(g,\pi)}$ (как и на любом другом J-инвариантном подпространстве многообразия $S_2 \times S_d^2$). Мы должны показать, что Ω' не вырождена. Предположим, что для $(h_1, \omega_1) \in \mathscr{G}_{(g,\pi)}$

$$\int_M \langle J^{-1}(h_1, \omega_1), (h_2, \omega_2)\rangle = 0$$

при всех $(h_2, \omega_2) \in \mathscr{G}_{(g,\pi)}$. Поскольку $\mathscr{G}_{(g,\pi)}$ J-инвариантно, то

$$J \circ \begin{pmatrix} h_1 \\ \omega_1 \end{pmatrix}^* \in \mathscr{G}_{(g,\pi)}.$$

Поэтому положим

$$\begin{pmatrix} h_2 \\ \omega_2 \end{pmatrix} = -J \circ \begin{pmatrix} h_1 \\ \omega_1 \end{pmatrix}^*$$

и, учитывая, что $J^* = -J$, получаем

$$0 = -\int \langle J^{-1}(h_1, \omega_1), J \circ (h_1, \omega_1)^*\rangle =$$

$$= \int \langle (h_1, \omega_1), (h_1, \omega_1)^*\rangle = \int_M (h_1 \cdot h_1 + \omega_1' \cdot \omega_1')\, d\mu(g)$$

Отсюда $(h_1, \omega_1) = 0$, и, следовательно, форма Ω' не вырождена.

Предложение 38 является частным случаем следующего общего результата симплектической геометрии [173].

Теорема 39

Пусть (V, Ω) — (слабо) симплектическое векторное пространство, а $W \subset V$ — некоторое подпространство. Пусть $W_\Omega^\perp = \{v \in V \mid \Omega(v, \omega) = 0$ для всех $\omega \in W\}$; предположим также, что W коизотропно, т. е. $W_\Omega^\perp \subset W$. Тогда $W/W_\Omega^\perp$ является (слабо) симплектическим векторным пространством на естественном пути.

Доказательство. Обозначим элемент из $W/W_\Omega^\perp$ через $w + W_\Omega^\perp$. Введем Ω на этом фактор-пространстве с помощью соотношения $\Omega(w_1 + W_\Omega^\perp, w_2 + W_\Omega^\perp) = \Omega(w_1, w_2)$. Поскольку $\Omega(w_i, W_\Omega^\perp) = 0$, $i = 1, 2$, это определение корректно. С другой стороны, если $\Omega(w_1 + W_\Omega^\perp, w_2 + W_\Omega^\perp) = 0$ для всех $w_1, w_2 \in W_\Omega^\perp$, то $w_2 + W_\Omega^\perp$ является нулевым элементом фактор-пространства.

■

Предложение 38 следует отсюда как частный случай, если поло-

жить $V=S_2\times S_d^2=T_{(g,\pi)}T^*\mathcal{M}$, взять Ω, как определено выше, и $W=\ker D\Phi(g, \pi)$. Тогда

$$\begin{aligned} W_\Omega^\perp &= \{(h, \omega) \mid \Omega((h, \omega), W)=0\} = \\ &= \{(h, \omega) \mid J^{-1}(h, \omega) \text{ ортогонально к } \ker D\Phi(g, \pi)\} = \\ &= \{(h, \omega) \mid J^{-1}(h, \omega) \in \operatorname{im}[D\Phi(g, \pi)]^*\} = \\ &= \operatorname{im}\{J\circ[D\Phi(g, \pi)]^*\} \subset W. \end{aligned}$$

■

Описанная выше симплектическая структура на $\mathcal{G}$ может оказаться важной для проблемы квантования гравитации. Эта симплектическая структура, вероятно, содержится в неявной форме в работах Бергмана [17], Дирака [72] и Де Витта [66]. Однако изложенная нами формулировка позволяет развить более геометричный и более строгий подход. Прежде всего он дает возможность воспользоваться формализмом Сигала или Костанта—Сурио для полного или полуклассического квантования. Во-вторых, подход, предлагаемый нами, позволяет показать, что вблизи метрики ${}^{(4)}g$ в $\mathcal{E}(V_4)$, не допускающей никакой изометрии (и, следовательно, никакого пространственно-временно́го векторного поля Киллинга), $\mathcal{G}=\mathcal{E}(V_4)/\mathcal{D}(V_4)$ является гладким многообразием, находится в естественном локальном изоморфизме фактор-пространству $\mathcal{C}_{\mathcal{H}}\cap\mathcal{C}_\delta/\!\sim$ и, следовательно, несет каноническую симплектическую структуру [1]). Таким образом, в окрестности пространств-времен с нулевым тензором Эйнштейна, не допускающих векторных полей Киллинга, пространство $\mathcal{G}=\mathcal{E}(V_4)/\mathcal{D}(V_4)$ гравитационных степеней свободы само является симплектическим многообразием, или, если угодно, гравитационным *фазовым пространством* без сингулярностей, каждый элемент которого представляет собой геометрию некоторого пустого пространства. Заметим, что $\mathcal{G}$ представляет собой (вообще говоря) симплектическое многообразие, несмотря на то, что не является кокасательным расслоением. Мы предполагаем, что $\mathcal{G}$ может быть в действительности разбито на симплектические многообразия, подобно стратификации суперпространства; см. [18, 82]. Сингулярности в $\mathcal{G}$ возникают в окрестности пространств-времен с симметриями, и они имеют конический характер [95]. В своем отчете 1978 г. Фонду гравитационных исследований Монкри подчеркивал, что эти сингулярности оказывают существенное влияние на процедуры квантования (например, для пространства-времени де Ситтера).

Методы, использованные нами для анализа гравитации, основаны на L_2-сопряженном формализме и переносятся непосредствен-

[1]) Интересным моментом является здесь то, что $\mathcal{C}_{\mathcal{H}}\cap\mathcal{C}_\delta$ хотя и представляет собой подмногообразие многообразия $T^*\mathcal{M}$, но не обладает естественной симплектической структурой, индуцированной из $T^*\mathcal{M}$, поскольку касательное пространство к $\mathcal{C}_{\mathcal{H}}\cap\mathcal{C}_\delta$ не J-инвариантно. Приходится переходить к фактор-многообразию $\mathcal{C}_{\mathcal{H}}\cap\mathcal{C}_\delta/\!\sim$, чтобы получить симплектическую структуру, индуцированную из $T^*\mathcal{M}$.

но на поля, минимально взаимодействующие с гравитацией, в частности на поля Янга—Миллса. В последнем случае эквивалентность устанавливается не по группе $\mathscr{D}(V_4)$, а по большей группе диффеоморфизмов эквивалентного расслоения (т. е. калибровочных преобразований, накрывающих диффеоморфизмы пространства-времени). С помощью изложенных методов мы можем показать, что пространство степеней свободы для полей и гравитации, если взаимодействие полей с гравитацией минимально, представляет собой, вообще говоря, симплектическое многообразие; см. [93, 94].

В заключение мы хотели бы выразить надежду, что представленные здесь методы помогут раскрыть некоторые взаимосвязи, существующие между общей теорией относительности, дифференциальной геометрией, функциональным анализом, теорией нелинейных уравнений в частных производных, бесконечномерными динамическими системами, симплектической геометрией и теорией сингулярностей. Несомненно, все эти (как и прочие) области математики должны внести соответствующий вклад в исследование теории гравитации, прежде чем наступит время окончательного анализа.

ЛИТЕРАТУРА [1])

1. *Abraham R., Marsden J.* Foundations of Mechanics, Second Edition. W. A. Benjamin, New York, 1978.
2. *Arms J.*, Linearization stability of the Einstein—Maxwell system, J. Math. Phys., **18**, 830 (1977).
3. *Arms J.*, Linearization stability of coupled gravitational and gauge fields. Thesis, Berkeley, 1977.
4. *Arms J., Fischer A., Marsden J.*, Une approche symplectique pour des théorèmes de décomposition en géometrie ou relativité générale, C. R. Acad. Sci. (Paris), **281**, 571 (1975).
5. *Arnowitt R., Deser S., Misner C. W.*, Dynamical structure and definition of energy in general relatiivty. Phys. Rev., **116**, 1322 (1959).
6. *Arnowitt R., Deser S., Misner C. W.*, Note on positive-definiteness of the energy of the gravitational field. Ann. Phys., **11**, 116 (1960).
7. *Arnowitt R., Deser S., Misner C. W.*, The dynamics of general relativity.— In: Gravitation: an Introduction to Current Research, ed. L. Witten, Wiley, New York, 1962, p. 227.
8. *Avez A.*, Essais de géométrie riemannienne hyperbolique global — applications a la relativité générale. Ann. Inst. Fourier (Grenoble), **13**, 105 (1963).
9. *Avez A.*, Le probleme des conditions initiales. In Fluides et champs gravitationnels en relativité générale, Colloq. Intern. CRNS 170. Paris, 1967, p. 163.
10. *Bancel D., Choquet-Bruhat Y.*, Existence, uniqueness and local stability for the Einstein—Maxwell—Boltzmann system. Commun. Math. Phys., **33**, 83 (1973).
11. *Bancel D., Lacaze J.*, Espaces de fonctions avec conditions asymptotiques et sections d'espace maximal non compact. C. R. Acad. Sci. (Paris), **283**, 405 (1976).
12. *Barbance C.*, Decomposition d'un tenseur symmetrique sur en espace d'Einstein. C. R. Acad. Sci. (Paris), **258**, 5336; **264**, 515 (1964).

[1]) Назначение этого списка литературы — дать достаточно полное представление о состоянии проблем, обсуждаемых в статье, поэтому отнюдь не на все работы, упомянутые здесь, имеется ссылка в тексте статьи.

13. *Berezdivin J.*, The analytic noncharacteristic Cauchy problem for nonlightlike isometrics in vacuum spacetimes. J. Math. Phys., **15**, 1963 (1974).
14. *Berger M.*, Sur les varietes d'Einstein compactes. In Comptes Rendu; III Reunion Math. Expression Latine Namur. Centre Belge de Recherches Mathématiques, 1965.
15. *Berger M.*, Quelques formules de variation pour une structure riemannienne, Ann. Scient. Ec. Norm. Sup., **3**, 285 (1970).
16. *Berger M.*, *Ebin D.*, Some decompositions of the space of symmetric tensors on a Riemannian manifold. J. Differ. Geom., **3**, 379 (1969).
17. *Bergmann P. G.*, Conservation Laws in general relativity as the generators of coordinate transformations. Phys. Rev., **112**, 287 (1958).
18. *Bourguignon J. P.*, Une stratification de l'espace des structures riemanniennes, Comp. Math., **30**, 1 (1975).
19. *Bourguignon J. P.*, *Ebin D.*, *Marsden J.*, Sur le noyau des operateurs pseudo-differentiels a symbole surjective et non-injective. C. R. Acad. Sci. (Paris), **282**, 867 (1976).
20. *Brezis H.*, Operateurs maximaux monotones. North-Holland, Amsterdam, 1973.
21. *Brill D.*, Maximal surfaces in closed and open spacetimes. Proceedings of the Marcel Grossman Meeting, July 1975, ed. R. Ruffini. North-Holland, Amsterdam, 1977.
22. *Brill D.*, *Deser S.*, Variational methods and positive energy in general relativity. Ann. Phys. (N. Y.), **50**, 548 (1968). [См. также: *Brill*, *Deser*, *Fadeev*, Phys. Lett., **26A**, 538 (1968).]
23. *Brill D.*, *Deser S.*, Instability of closed spaces in general relativity. Commun. Math. Phys., **32**, 291 (1973).
24. *Budic R.*, *Isenberg J.*, *Lindblom L.*, *Yasskin P.*, On the determination of Cauchy surfaces from intrinsic properties. Commun. Math. Phys., **61**, 87 (1978).
25. *Cantor M.*, Spaces of functions with asymptotic conditions on R^n. Indiana University Math. J., **24**, 897 (1975).
26. *Cantor M.*, The existence of non-trivial asymptotically flat initial data for vacuum spacetimes. Commun. Math. Phys., **57**, 83 (1977).
27. *Cantor M.*, Some problems of global analysis on asymptotically simple manifolds. Comp. Math., 1978.
28. *Cantor M.*, *Fischer A.*, *Marsden J.*, *O'Murchadha N.*, *York J.*, On the existence of maximal slices. Commun. Math. Phys., **49**, 897 (1976).
29. *Caricato C.*, Sur le probléme de Cauchy intrinsèque pour les equations de Maxwell—Einstein dans le vide. Ann. Inst. H. Poincare, Sect. A, **11**, 373 (1969).
30. *Chernoff P.*, *Marsden J.*, Properties of Infinite-Dimensional Hamiltonian Systems. Springer Lecture Notes No. 425, 1974.
31. Choquet (Foures)-Bruhat Y. Sur l'intégration des équations d'Einstein. C. R. Acad. Sci. (Paris), **226**, 1071 (1948). [См. также: *Ration J.* Mech. Anal., **5**, 951 (1956).]
32. *Choquet-Bruhat Y.*, Théorème d'existence pour certains systémes d'équations aux derivées partielles non linéaires. Acta Math., **88**, 141 (1952).
33. *Choquet-Bruhat Y.*, Théorèmes d'existence an mécanique des fluides rélativistes. Bull. Soc. Math. France, **86**, 155 (1958).
34. *Choquet-Bruhat Y.*, Fluides rélativistes de conductivité infinie. Astron. Acta, **6**, 354 (1960).
35. *Choquet-Bruhat Y.*, Cauchy problem. In: Gravitation: an Introduction to Current Research, ed. L. Witten. Wiley, New York, 1962.
36. *Choquet-Bruhat Y.*, Espaces-temps einsteiniens généraux choes gravitationels. Ann. Inst H. Poincaré, **8**, 327 (1968).
37. *Choquet-Bruhat Y.*, Etude des equations des fluides chargés rélativistes inductifs et conducteurs. Commun. Math. Phys., **3**, 334 (1968).
38. *Choquet-Bruhat Y.*, Mathematicals problems in general relativity. Proc. Int. Congress, Nice. French Mathematics Society, 1970.
39. *Choquet-Bruhat Y.*, New elliptic system and global solutions for the constraints equations in general relativity. Commun. Math. Phys., **21**, 211 (1971).

40. *Choquet-Bruhat Y.*, Solutions C^∞ d'equations hyperbolique non linéaires. C. R. Acad. Sci. (Paris), **272**, 386 (1971).
41. *Choquet-Bruhat Y.*, Probleme de Cauchy pour le systéme integrodifferentiel d'Einstein—Liouville. Ann. Inst. Fourier XXI, **3**, 181 (1971).
42. *Choquet-Bruhat Y.*, Stabilité de solutions d'équations hyperboliques non linéaires. Application a l'espace-temps de Minkowski en rélativité génerale. C. R. Acad. Sci. (Paris), **274**, Ser. A, 843 (1972). [См. также: УМН, XXIX (2), 176, 314.]
43. *Choquet-Bruhat Y.*, C^∞ solutions of hyperbolic non linear equations. Gen. Relativ. Gravit., **2**, 359 (1972).
44. *Choquet-Bruhat Y.*, Global solutions of the equations of constraints in general relativity on closed manifolds. Symposia Math., **XII**, 317 (1973).
45. *Choquet-Bruhat Y.*, Sous-variétés maximales, ou a courbure constante, de variétés lorentziennes. C. R. Acad. Sci. (Paris), **280**, 169 (1975).
46. *Choquet-Bruhat Y.*, Quelques proprietés des sous-variétés maximales d'un variété lorentzienne. C. R. Acad. Sci. (Paris), **281**, 577 (1975).
47. *Choquet-Bruhat Y.*, The problem of constraints in general relativity, solution of the Lichnerowicz equation. In: Differential Geometry and Relativity, eds. M. Cahen and M. Flato. Riedel, Dordrecht, 1976.
48. *Choquet-Bruhat Y.*, Maximal submanifolds and manifolds with constant mean, extrinsic curvature of a Lorentzian manifold. Ann. Scuola Norm. Pisa, Serie IV, vol. III (in honour of J. Leray), 1976, p. 361.
49. *Choquet-Bruhat Y.*, *Deser S.*, Stabilité initiale de l'espace temps de Minkowski. C. R. Acad. Sci. (Paris), **275**, 1019 (1972).
50. *Choquet-Bruhat Y.*, *Geroch R.*, Global aspects of the Cauchy problem in general relativity. Commun. Math. Phys., **14**, 329 (1969).
51. *Choquet-Bruhat Y.*, *Lamoureau-Brousse L.*, Sur les equations de l'elasticite relativiste. C. R. Acad. Sci. (Paris), **276**, 1217 (1973).
52. *Choquet-Bruhat Y.*, *Marsden J.*, Solution of the local mass problem in general relativity. C. R. Acad. Sci. (Paris), **282**, 609; Commun. Math. Phys., **51**, 283 (1976).
53. *Choquet-Bruhat Y.*, *York J. W.*, In Einstein Centenary Volume, eds. P. Bergmann, J. Goldberg and A. Held, 1979.
54. *Choquet-Bruhat Y.*, *Fischer A.*, *Marsden J.*, Maximal hypersurfaces and positivity of mass. Proc. Enrico Fermi Summer School of the Italian Physical Society Varenna, ed. J. Ehlers, 1978.
55. *Christodoulou D.*, *Francaviglia F.*, The thin sandwich conjecture. Proc. Enrico Fermi Summer School Varenna, ed. J. Ehlers, 1978.
56. *Coll B.*, Sur la détermination, par des donnés de Cauchy, des champs de Killing admis par un espace-temps, d'Einstein—Maxwell. C. R. Acad. Sci. (Paris), **281**, 1109 (1975).
57. *Coll B.*, Sur la stabilité linéaire des équations d'einstein du vide. C. R. Acad. Sci. (Paris), **282**, 247 (1976).
58. *Coll B.*, On the evolution equations for Killing fields. J. Math. Phys., **18**, 1918 (1978).
59. *Cordero P.*, *Teitelboim C.*, Hamiltonian treatment of the spherically symmetric Einstein—Yang—Mills system. Ann. Phys. (N. Y.), **100**, 607 (1976).
60. *Courant R.*, *Hilbert D.*, Methods of Mathematical Physics, vol. II. Interscience, New York, 1962. [Имеется перевод: *Курант Р.*, *Гильберт Д.* Методы математической физики. Т. 2.— М.— Л.: Гостехиздат, 1965.]
61. *D'Eath P. D.*, On the existence of perturbed Robertson—Walker universes. Ann. Phys. (N. Y.), **98**, 237 (1976).
62. *Deser S.*, Covariant decomposition of symmetric tensors and the gravitational Cauchy problem. Ann. Inst. H. Poincare, **7**, 149 (1967).
63. *Deser S.*, Gravitational energy-momentum on nonmaximal surfaces. Phys. Rev., **D12**, 943 (1975).
64. *Deser S.*, *Teitelboim C.*, Supergravity has positive energy. Phys. Rev. Lett., **39**, 249 (1976).

65. *DeWitt B.*, Quantum theory of gravity, I. Canonical Theory. Phys. Rev., **160**, 1113 (1967).
66. *DeWitt B.*, Covariant quantum geometrodynamics. In: Magic Without Magic, ed. J. R. Klauder. W. H. Freeman. San Francisco, 1972.
67. *Dionne P.*, Sur les problemes de Cauchy bien posés. J. Anal. Math. Jerusalem, **10**, 1 (1962).
68. *Dirac P. A. M.*, Generalized Hamiltonian dynamics. Can. J. Math., **2**, 129 (1950).
69. *Dirac P. A. M.*, The Hamiltonian form of fields dynamics, Can. J. Math., **3**, 1 (1950).
70. *Dirac P. A. M.*, Generalized Hamiltonian dynamics. Proc. R. Soc. Lond., **A246**, 326 (1958).
71. *Dirac P. A. M.*, The theory of gravitation in Hamiltonian form. Proc. R. Soc. Lond., **A246**, 333 (1958).
72. *Dirac P. A. M.*, Fixation of coordinates in the Hamiltonian theory of gravitation. Phys. Rev., **114**, 924 (1959).
73. *Dirac P. A. M.*, Lectures on Quantum Mechanics. Belfer Graduate School of Science, Monograph Series No. 2, Yeshiva University, New York, 1964. [Имеется перевод: *Дирак П. А. М.* Лекции по квантовой механике.— М.: Мир, 1968.]
74. *Dorroh J. R.*, *Marsden J. E.*, Differentiability of nonlinear semigroups, 1975 (неопубликованные заметки).
75. *Ebin D.*, On the space of Riemannian metrics. Symp. Pure Math., Am. Math. Soc., **15**, 11 (1970).
76. *Ebin D.*, *Marsden J.*, Groups of diffeomorphisms and the motion of an incompressible fluid. Ann. Math., **92**, 102 (1970).
77. *Eells J.*, *Sampson J.*, Harmonic maps of Riemannian manifolds. Am. J. Math., **86**, 109 (1964).
78. *Einstein A.*, Hamiltonsches Princip und allgemeine Relativitatstheorie. Sitzungsber. Preuss. Akad. Wiss., 1111 (1916).
79. *Einstein A.*, Naherungsweise Integration der Feldgleichungen der Gravitation. Sitzungsber. Preuss. Akad. Wiss., 688 (1916).
80. *Einstein A.*, Der Energiesatz in der allgemeinen Relativitatstheorie. Sitzungsber. Preuss. Akad. Wiss., 448 (1918).
81. *Fadeev L. D.*, Symplectic structure and quantization of the Einstein gravitation theory. Actes du Congres Intern. Math., **3**, 35 (1970).
82. *Fischer A.*, The theory of superspace. In: Relativity, eds. M. Carmelli, S. Fickler, L. Witten. Plenum Press, New York, 1970.
83. *Fischer A.*, *Marsden J.*, The Einstein equations of evolution—a geometric approach. J. Math. Phys., **13**, 546 (1972).
84. *Fischer A.*, *Marsden J.*, The Einstein evolution equations as a first-order symmetric hyperbolic quasilinear system. Commun. Math. Phys., **28**, 1 (1972).
85. *Fischer A.*, *Marsden J.*, Linearization stability of the Einstein equations. Bull. Am. Math. Soc., **79**, 995 (1973).
86. *Fischer A.*, *Marsden J.*, New theoretical techniques in the study of gravity. Gen. Relativ. Grav., **4**, 309 (1973).
87. *Fischer A.*, *Marsden J.*, Global analysis and general relativity. Gen. Relativ. Grav., **5**, 89 (1974).
88. *Fischer A.*, *Marsden J.*, Linearization stability of nonlinear partial differential equations. Proc. Symp. Pure Math. Am. Math. Soc., **27**, Part 2, 219 (1975).
89. *Fischer A.*, *Marsden J.*, Deformations of the scalar curvature. Duke Math. J., **42**, 519 (1975).
90. *Fischer A.*, *Marsden J.*, A new Hamiltonian structure for the dynamics of general relativity. Gen. Relativ. Grav., **12**, 915 (1976).
91. *Fischer A.*, *Marsden J.*, The manifold of conformally equivalent metrics. Can. J. Math., **29**, 193 (1977).
92. *Fischer A.*, *Marsden J.*, Topics in the dynamics of general relativity. In: Proceedings Enrico Fermi Summer School of the Italian Physical Society, Varenna, ed. J. Ehlers, 1978.

93. *Fischer A., Marsden J.*, Hamiltonian field theories on spacetime, 1978.
94. *Fischer A., Marsden J.*, The space of gravitational degrees of freedom, 1978.
95. *Fischer A., Marsden J., Moncrief V.*, The structure of the space solutions of Einstein's equations, 1978.
96. *Frankl F.*, Über das Angangswertproblem fur lineare und nichtlineare partielle Differentialgleichungen zweiter Ordnung. [См. также: Мат. сб., 1937, т. 2 (44), с. 814.]
97. *Friedman A.*, Partial gifferential equations. Holt, New York, 1969.
98. *Friedrichs K. O.*, Symmetric Hyperbolic linear differential equations. Commun. Pure Appl. Math., **8**, 345 (1954).
99. *Friedrichs K.*, On the laws of relativistic magneto-fluid dynamics. Commun. Pure Appl. Math., **27**, 749 (1974).
100. *Garding L.*, Cauchy's problem for hyperbolic equations. Lecture Notes. University of Chicago, 1957. [Имеется перевод: *Гординг Л.* Задача Коши для гиперболических уравнений.— М.: ИЛ, 1961.]
101. *Garding L.*, Energy inequalities for hyperbolic systems. In: Differential Analysis. Bombay Colloquium, Oxford University Press, 1964, p. 209.
102. *Geroch R.*, General relativity. Proc. Symp. Pure Math., vol. 27 (part 2), 1975, p. 401; J. Math. Phys., **13**, 956 (1972); Ann. N. Y. Scad. Sci., **224**, 108 (1973).
103. *Hanson A., Regge T., Teitelboim C.*, Constrained Hamiltonian systems. Accademia Nazionale dei Lincei, Rome, No. 22, 1976, p. 1.
104. *Hawking S. W., Ellis G. F. R.*, The Large Scale Structure of space-time. Cambridge University Press, 1973. [Имеется перевод: *Хокинг С., Эллис Дж.* Крупномасштабная структура пространства-времени.— М.: Мир, 1977.]
105. *Hermann R.*, Gauge Fields and Cartan—Ehresman Connections. Math-Sci Press, Brookline, Massachusetts, 1975.
106. *Hille E., Phillips R.*, Functional Analysis and Semigroups. Americal Mathematical Society. Providence, Rhode Island, 1967. [Имеется перевод: *Хилле Э., Филлипс Р.* Функциональный анализ и полугруппы.— М.: ИЛ, 1962.]
107. *Hormander L.*, Pseudo-differential operators and non-elliptic boundary value problems. Ann. Math., **83**, 129 (1966).
108. *Hugles T., Kato T., Marsden J.*, Well-posed quasi-linear second-order hyperbolic systems with applications to nonlinear elastodynamics and general relativity. Arch. Ration. Mech. Anal., **63**, 273 (1977).
109. *Kato T.*, Perturbation Theory for Linear Operators. Springer, Berlin, Heidelberg, New York, 1966. [Имеется перевод: *Като Т.* Теория возмущений линейных операторов.— М.: Мир, 1972.]
110. *Kato T.*, Linear evolution equations of «hyperbolic» type. J. Fac. Sci. Univ. Tokyo, **17**, 241 (1970).
111. *Kato T.*, Linear evolution equations of «hyperbolic» type. II Math. Soc. Japan, **25**, 648 (1973).
112. *Kato T.*, The Cauchy problem for quasi-linear symmetric hyperbolic systems. Arch. Ration. Mech. Anal., **58**, 181 (1975).
113. *Kato T.*, Quasi-linear equations of evolution with applications to partial differential equations. Springer Lecture Notes, **448**, 25 (1975).
114. *Kato T.*, Linear and Quasilinear Equations of Evolution of Hyperbolic Type. Bressanone Lectures. Centro Internazionale Matematico Estivo, Rome, 1977.
115. *Kazdan J., Warner F.*, A direct approach to the determination of Gaussian and scalar curvature functions. Invent. Math., **28**, 227 (1975).
116. *Kijowski J., Szczyrba W.*, A canonical structure for classical field theories. Commun. Math. Phys., **46**, 183 (1976).
117. *Kobayashi S.*, Transformation Groups in Differential Geometry. Springer, Berlin, Heidelberg, New York, 1974.
118. *Kryzanski M., Schauder J.*, Quasilineare Differentialgleichungen zweiter Ordnung von hyperbolischen Typus. Gemischte Randwertaufgaben, Studia Math., 162 (1934).
119. *Kuchar K.*, A bubble-time canonical formalism for geometrodynamics. J. Math. Phys., **13**, 768 (1972).

120. *Kuchař K.*, Geometrodynamics regained: a Lagrangian approach. J. Math. Phys., 15, 708 (1974).
121. *Kuchař K.*, Geometry of hyperspace. I. J. Math. Phys., 17, 777 (1976).
122. *Kuchař K.*, Kinematics of tensor fields in hyperspace. II. J. Math. Phys., 17, 792 (1976).
123. *Kuchař K.*, Dynamics of tensor fields in hyperspace. III. J. Math. Phys., 17, 801 (1976).
124. *Kuchař K.*, Geometrodynamics with tensor sources. IV. J. Math. Phys., 18, 1589 (1977).
125. *Kuchař K.*, On equivalence of parabolic and hyperbolic super-Hamiltonians. J. Math. Phys., 19, 390 (1978).
126. *Lanczos C.*, Ein verinfachendes Koordinatensystem für die Einsteinschen Gravitationsgleichungen. Phys. Z., 23, 537 (1922).
127. *Lax P.*, Cauchy's problem for hyperbolic equations and the differentiability of solutions of elliptic equations. Commun. Pure Appl. Math., 8, 615 (1955).
128. *Leray J.*, Hyperbolic Differential Equations. Institute for Advanced Study (Notes), 1953.
129. *Lichnerowicz A.*, Problèmes globaux en mécanique relativiste. Hermann, 1939.
130. *Lichnerowicz A.*, L'intégration des équations de la gravitation problème des *n* corps. J. Math. Pures. Appl., 23, 37 (1944).
131. *Lichnerowicz A.*, Théories relativistes de la gravitation et de l'électromagnetism. Masson, Paris, 1955.
132. *Lichnerowicz A.*, Propagateurs et commutateurs en relativité générale. Publications Mathématiques No. 10, p. 293. Institutes des Hautes Études Scientifiques, Paris, 1961.
133. *Lichnerowicz A.*, Relativistic Hydrodynamics and Magnetohydramics: Lectures on the Existence of Solutions. W. A. Benjamin, New York, 1967.
134. *Lichnerowicz A.*, Ondes de choc et hypothèses de compressibilité en MHD rélativiste. Commun. Math. Phys., 12, 145 (1969).
135. *Lichnerowicz A.*, Ondes de choc gravitationnelles et électromagnetiques. Symposia Math. Inst. Naz. di Alta Matematica, Rome, 12, 93 (1972).
136. *Lichnerowicz A.*, Shock waves in relativistic magnetohydrodynamics under general assumptions. J. Math. Phys., 17, 2135 (1976).
137. *Lions J. L.*, Quelques methodes de résolution des problèmes nonlineares, Dunod, Paris, 1969.
138. *Marsden J.*, Applications of Global Analysis in Mathematical Physics. Publish or Perish, Boston, 1974.
139. *Marsden J.*, *Fischer A.*, On the existence of nontrivial, complete, asymptotically flat spacetimes. Publ. Univ. Lyon, 4 (fasc. 2), 182 (1972).
140. *Marsden J.*, *Hughes T.*, Topics in Mathematical Foundations of Elasticity. In Non-linear Analysis and Mechanics, vol. II, ed. R. J. Knops, Pitman, 1978.
141. *Marsden J. E.*, *Tipler F. J.*, Maximal hypersurfaces and foliations of constant mean curvature in general relativity. (Preprint).
142. *Marsden J.*, *Weinstein A.* Reduction of symplectic manifolds with symmetry. Rep. Math. Phys., 5, 121 (1974).
143. *Marsden J.*, *Ebin D.*, *Fischer A.*, Diffeomorphism groups, hydrodynamics, and general relativity. Proc. 13th Beinniel Seminar of the Canadian Math. Cong. Montreal, 135 (1972).
144. *Misner C. W.*, *Thorne K.*, *Wheeler J. A.*, Gravitation W. H. Freeman, San Francisco, 1974. [Имеется перевод: *Мизнер Ч.*, *Торн К.*, *Уилер Дж.* Гравитация.— М.: Мир, 1977.]
145. *Moncrief V.*, Spacetime symmetries and linearization stability of the Einstein equations. I. J. Math. Phys., 16, 493 (1975).
146. *Moncrief V.*, Decompositions of gravitational perturbations. J. Math. Phys., 16, 1556 (1975).
147. *Moncrief V.*, Gauge-invariant perturbations of Reissner—Nordstrom black holes. Phys. Rev., D12, 1526 (1975).

148. *Moncrief V.*, Spacetime symmetries and linearization stability of the Einsteni equations. II. J. Math. Phys., **17**, 1893 (1976).
149. *Moncrief V.*, Gauge symmetries of Yang—Mills fields. Ann. Phys. (N. Y.), **108**, 387 (1977).
150. *Müller zum Hagen H.*, *Seifert H. J.*, On characteristic initial-value and mixed problems. Gen. Relativ. Grav., 1976.
151. *Nijenhuis A.*, X_{n-1} forming sets of eigenvectors. Proc. Kon. Ned. Akad. Amsterdam, **54**, 200 (1951).
152. *Nirenberg L.*, On elliptic partial differential equations. Ann. Scuola. Norm. Sup. Pisa, **13**, 115 (1959).
153. *Nirenberg L.*, *Walker H.*, The null space of elliptic partial differential operators in R^n. J. Math. Anal. Appl., **47**, 271 (1973).
154. *O'Murchadha N.*, *York J. W.*, The initial-value problem of general relativity. Phys. Rev., **D10**, 428 (1974).
155. *O'Murchadha N.*, Gravitational energy. Phys. Rev., **D10**, 2345 (1974).
156. *Palais R.*, Natural operations on differential forms. Trans. Am. Math. Soc., **92**, 125 (1959).
157. *Palais R.*, Seminar on the Atiyah-Singer Index Theorem. Princeton Univeristy Press, 1965.
158. *Palais R.*, Foundations of Global Nonlinear Analysis. Benjamin, New York, 1968.
159. *Петровский И. Н.* Мат. сб., 1937, т. 2, с. 814.
160. *Pham Mau Quan.*, Magneto-hydrodynamique relativiste. Ann. Inst. H. Poincaré. Nouv. Serie t. 2, 151 (1965).
161. *Regge T.*, *Teitelboim C.*, Role of surface integrals in the Hamiltonian formulation of general relativity. Ann. Phys. (N. Y.), **88**, 286 (1974).
162. *Rund H.*, *Lovelock D.*, Variational principles in the general theory of relativity. Über Deutsch. Math-Verein., **74**, 1 (1972).
163. *Sachs R.*, *Wu H.*, General Relativity and Cosmology for Mathematicians. Springer-Verlag Graduate Texts in Mathematics. Springer, Berlin, Heidelberg, New York, 1977.
164. *Schauder J.*, Das Anfangswertproblem einer quasi-linearen hyperbolischen Differentialgleichung zweiter Ordnung in beliebiger Anzahl von unabhangigen Veranderlichen. Fund. Math., **24**, 213 (1935).
165. *Schwartz J.*, Nonlinear Functional Analysis. Gordon and Breach, New York, 1969.
166. *Szczryba W.*, On geometric structure of the set of solutions of Einstein equations. Diss. Math., **150**, 1 (1977).
167. *Segal I.*, Symplectic structures and the quantization problem for wave equations. Symp. Math., **14**, 99 (1974).
168. *Соболев С. Н.* Мат. сб., 1939, т. 5, с. 71.
169. *Соболев С. Н.* Некоторые применения функционального анализа в математической физике.— Новосибирск: Изд. сиб. отд. АН СССР, 1962.
170. *Souriau J. M.*, Structure des systemes dynamique. Dunod, Paris, 1970.
171. *Taub A.*, Variational Principles in General Relativity, Bressanone Lectures, Centro Internazionale Matematico Estivo, Rome, 1970, p. 206.
172. *Terng C. L.*, Natural Vector Bundles and Natural Differential Operators. Thesis, Brandeis University, Walthamn, Massachusetts, 1976.
173. *Weinstein A.*, Lectures on Sympletic Manifolds. CBMS No. 29. American Mathematical Society, Providence, Rhode Island, 1977.
174. *Wheeler J. A.*, Geometrodynamics and the issue of the final state. In: Relativity, Groups, and Topology, eds. C. M. DeWitt, B. S. DeWitt. Gordon and Breach, New York, 1964.
175. *Wheeler J. A.*, Geometrodynamics. Academic Press, New York, London, 1962.
176. *York J. W.*, Gravitational degrees of freedom and the initial-value problem. Phys. Rev. Lett., **26**, 1656 (1971).
177. *York J. W.*, Mapping onto solutions of the gravitational initial-value problem. J. Math. Phys., **13**, 125 (1972).

178. *York J. W.*, Role of conformal three-geometry in the dynamics of gravitation. Phys. Rev. Lett., **28**, 1082 (1972).
179. *York J. W.*, Conformally invariant orthogonal decomposition of symmetric tensors on Riemannian manifolds and the initial-value problem of general relativity. J. Math. Phys., **14**, 456 (1973).
180. *York J. W.*, Covariant decompositions of symmetric tensors in the gravitation. Ann. Inst. H. Poincaré, **21**, 319 (1974).
181. *Yosida K.*, Functional Analysis, Fourth edition, Springer, Berlin, Heidelberg, New York, 1974. [Имеется перевод: *Иосида К.* Функциональный анализ.— М.: Мир, 1967.]

III. АСТРОФИЗИКА ЧЕРНЫХ ДЫР

Р. Д. Блэндфорд [1]), К. С. Торн [2]) [3])

1. ВВЕДЕНИЕ

1.1. НЕМНОГО ИСТОРИИ

Хотя первые теоретические работы, посвященные черным дырам [44, 107, 137], были мотивированы астрофизическими соображениями, до самого недавнего времени астрономы-практики не принимали черных дыр всерьез. Один из авторов (К. С. Т.) помнит предупреждение, которое он слышал, когда заканчивал Калифорнийский технологический институт в конце 50-х годов,— что ОТО, вероятнее всего, не нужна в астрономии, за исключением ситуации «большого взрыва». «Все звезды, по-видимому, теряют достаточно много массы в процессе своей эволюции,— говорил профессор астрономии,— чтобы оказаться ниже чандрасекаровского предела и тем самым избежать гравитационного коллапса — судьбы, предсказанной в работе Оппенгеймера».

Такая позиция господствовала среди астрономов до тех пор, пока два открытия не потрясли оснований астрономии. Первым было открытие квазаров [177] с их гигантским энерговыделением, которое наводило на мысль о гравитационном коллапсе как движущем механизме. Вторым было открытие пульсаров [81] и приведенные год спустя свидетельства в пользу того, что их энергетика и пульсации излучения обеспечиваются вращающимися нейтронными звездами, которые, по-видимому, образовались при взрывах Сверхновых [49, 69, 146]. Реакция астрономов на эти открытия характеризуется случаем с одним известным астрофизиком, который в начале 60-х годов спрашивал: «Сколько ангелов могут сегодня танцевать на макушке нейтронной звезды?» и который в конце 60-х и начале 70-х годов приступил к астрофизическому исследованию не только нейтронных звезд, но также и черных дыр.

Некоторые (но не все!) астрономы так хотят теперь верить в черные дыры, что только и слышно, как черные дыры предлагаются в качестве deus ex machina [4]) почти в каждом случае открытия нового

---

[1]) *R. D. Blandford*, Department of Theoretical Astrophysics, California Institute of Technology, Pasadena, California, USA.

[2]) *K. S. Thorne*, Department of Theoretical Astrophysics, California Institute of Technology, Pasadena, California, USA.

[3]) Работа частично субсидирована Национальным научным фондом [AST76-20375 и AST76-80801 A01].

[4]) «Бог из машины» (*лат.*). В конце античных пьес бог появляется на сцене как средство привести сюжет к развязке.— *Прим. перев.*

таинственного наблюдаемого явления — от двойного рентгеновского источника Cygnus X-1 и других рентгеновских источников [28, 215] до ситуации с нехваткой солнечных нейтрино [48, 193], до загадки недостающей массы, требуемой для стабилизации спиральных галактик [91, 141], для связывания скоплений галактик (см., например, [153]) или даже замыкания Вселенной (например, [38, 202]), до падения в прошлом веке Тунгусского метеорита [92] и до исчезновения судов в Бермудском треугольнике (см., например, [19]). Дело приняло столь одиозный характер, что возникла насущная необходимость отделить правдоподобные гипотезы, касающиеся роли черных дыр, от неправдоподобных.

Сейчас в астрофизике черных дыр имеется достаточно тесное взаимодействие между теорией и наблюдениями. Это не всегда было так. Пионерские работы Лапласа [107] и Оппенгеймера и Снайдера [137] были очень мало стимулированы наблюдательной астрономией. Впоследствии, в конце 50-х и начале 60-х годов, без каких-либо наблюдательных данных Уилер [219, 220] побуждал своих коллег-физиков рассматривать гравитационный коллапс и возникающую сингулярность как «один из величайших кризисов всех времен» для фундаментальной физики. Возможно, именно благодаря влиянию Уилера астрофизики, когда были открыты квазары, сразу стали искать объяснение в гравитационном коллапсе. И в последующем интеллектуальном брожении астрофизики начали спрашивать себя, каких других наблюдательных проявлений можно было бы ожидать от сколлапсировавших объектов. С этого и началась современная эра астрофизики черных дыр с ее наблюдательно-теоретическими связями, хотя термин «черная дыра» был придуман еще Уилером [221].

1.2. О ХАРАКТЕРЕ ЭТОЙ СТАТЬИ

В последние годы ряд талантливых физиков-теоретиков успешно переходит от общей теории относительности и квантовой теории поля к исследованию астрофизических проблем, включающих черные дыры. Этот переход часто чрезвычайно труден. Он влечет за собой освоение большого объема новой физики, а также, что более важно, освоение совершенно нового стиля исследований.

Эта статья адресована тем физикам-теоретикам, которые поддаются искушению вслед за своими коллегами заняться астрофизикой, тем, кто, не стремясь к переменам, хочет посмотреть на мир с других позиций, студентам, которые могли бы захотеть заняться астрофизикой, но до сих пор мало сталкивались с ней, и, наконец, тем, кому просто интересно узнать, какую роль могли бы играть черные дыры в астрофизической картине мира. Эта статья не адресована уже работающим астрофизикам.

Вследствие такой направленности (а также по лености) мы воздержались от проведения тщательного исследования литературы по

астрофизике черных дыр и от написания ее подробного обзора. (В дальнейшем будет очевидно, что мы в долгу перед более ранними обзорами, например [59].) Вместо этого мы делаем следующее. В разд. 2 мы пытаемся дать читателю общее представление о характере исследований в астрофизике черных дыр, т. е. о знаниях, необходимых для работы, и о стиле, в котором думают и вычисляют. Затем в разд. 3—7 мы переходим к обзору наиболее правдоподобных современных сценариев, касающихся роли черных дыр во вселенной, и наблюдательных данных, на которых проверяются эти сценарии. Для каждого сценария дается только несколько ссылок: те, которые, как мы думаем, могли бы быть наиболее полезны для новичка в этой области, или те, которые могли бы указать ему наиболее прямой путь в литературу. Зачастую это означает, что мы ссылаемся скорее на обзоры, чем на оригинальные статьи. Наши сценарии организованы в разделы следующим образом:

3. Черные дыры «звездной массы» ($M \sim (1—100) M_\odot$), одиночные в межзвездной среде.
4. Черные дыры звездной массы в двойных системах.
5. Черные дыры, которые могли бы образоваться и существовать в шаровых скоплениях.
6. Черные дыры в ядрах галактик и квазарах.
7. Черные дыры, которые могли бы образоваться до галактик, либо в само́м «большом взрыве», либо в догалактических конденсациях.

1.3. НЕКОТОРЫЕ МОЛЧАЛИВО ПОДРАЗУМЕВАЕМЫЕ ПРЕДПОЛОЖЕНИЯ АСТРОФИЗИКИ ЧЕРНЫХ ДЫР

В литературе и в процессе работы в этой области с самого начала принимается несколько недоказанных предположений, которые обычно явно не формулируются. Они включают в себя:

1. «Предположение обруча» — что черная дыра образуется тогда и только тогда, когда масса M становится компактной в области, окружность которой в *любом* направлении есть $\mathscr{C} \leqslant 4\pi M$, так что обруч такого размера может скользить по этой области и поворачиваться на 360° (см. дополнение 32.3 в книге Мизнера, Торна и Уилера [133], в дальнейшем цитируемой как МТУ).

2. «Гипотезу космической цензуры» — в процессе превращения компактного объекта в черную дыру и впоследствии в ходе эволюции дыр в реальной вселенной никогда не образуются «голые» сингулярности (см. Пенроуз, [155]), за исключением, может быть, конечного момента испарения черной дыры.

3. «Предположение о быстрой потере волос» — что в отсутствие сильных внешних динамических гравитационных возмущений черная дыра с массой M будет приходить в стационарное состояние за время

$$\Delta t \sim 100M \sim 10^{-3} \text{ с} \, (M/M_\odot)$$

(см., например, дополнение 32.2 в книге МТУ [133], распространенное на характерные ситуации).

4. «Гипотезу слияния» — что при столкновении двух черных дыр они всегда сливаются, образуя единую черную дыру (см., например, [189] в применении к характерным ситуациям).

Кроме того, астрофизические исследования, безусловно, опираются на важные основные теоремы и результаты, полученные в физике черных дыр.

2. О ХАРАКТЕРЕ ИССЛЕДОВАНИЙ В АСТРОФИЗИКЕ ЧЕРНЫХ ДЫР

Фундаментальная теория черных дыр является хорошо обоснованной, элегантной, чистой и самосогласованной. Она неумолимо и определенно следует из фундаментальных законов физики. Теория черных дыр в астрофизическом окружении представляет собой полную противоположность. Поскольку она имеет дело с физикой больших объемов материи — обращающейся вокруг дыры и аккрецирующей на дыру,— она подвержена всем запутанным неясностям и неопределенностям современной теории поведения больших количеств материи. Если на Земле грозы и ураганы не поддаются точному теоретическому моделированию, то можно ли надеяться предсказать хотя бы качественно поведение их аналогов в турбулентной, замагниченной плазме, которая аккрецирует на черную дыру в двойной системе? Нельзя. Самое лучшее, чего можно ожидать,— это получение очень грубых моделей буйного поведения материи в окрестности черной дыры.

К счастью, однако, окончательные модели имеют некую скромную надежду походить на реальность. Это происходит потому, что относительная важность физических процессов вблизи дыры может быть охарактеризована безразмерными отношениями, которые обычно оказываются очень большими, откуда следует, что буйное поведение материи вблизи дыры определяется небольшим числом процессов. Задача автора модели заключается в идентификации доминирующих процессов в заданной ему ситуации и в получении приближенных уравнений, описывающих их макроскопические проявления. История показывает, что идентифицировать главные процессы нелегко. Это связано с многочисленностью возможных процессов, подлежащих рассмотрению, и часто автор модели, даже вполне компетентный, может пропустить важный процесс.

Таким образом, исследования в области астрофизики черных дыр опираются на многие разделы физической теории. В каждом таком разделе надо как свои пять пальцев знать приближенные формулы, характеризующие длинный список возможных, относящихся к делу процессов. Необходимые области включают ОТО, физику равновесной и неравновесной плазмы, физику радиационных процессов и физику динамических звездных систем. Те характерные процессы,

которые были идентифицированы как важные для астрофизики черных дыр к 1972 г., собраны в обзоре Новикова и Торна [135]. После 1972 г. теоретики, рассматривающие возможные сценарии с участием черных дыр, столкнулись с новыми важными процессами, включающими: 1) возможность различной электронной и протонной температуры в плазме вокруг черной дыры (фундаментальная теория изложена Спитцером [190], приложения — Шапиро, Лайтманом и Эрдли [182]); 2) электродинамические процессы вблизи дыр [26] и 3) звезднодинамические процессы в окрестности черных дыр (см. обзор Лайтмана и Шапиро [114]).

Исследования по астрофизике черных дыр требуют также хорошего знания феноменологии современной астрономии — наблюдаемых свойств звезд, основных особенностей их эволюции, структуры Галактики и наблюдаемых физических условий в межзвездном пространстве. Возможный путь изучения этих вопросов — чтение соответствующих разделов хорошего элементарного учебника по астрономии (например, [1, 93, 126, 152]) и вслед за этим просматривание астрономических статей за последние 10 лет в журнале Scientific American. Только после этого читатель реально готов воспринимать обзорные статьи из журнала Annual Reviews of Astronomy and Astrophysics [1]).

Стиль работы в астрофизике черных дыр сильно отличается от стиля работы в ОТО или других областях фундаментальной теории. Грубо говоря, этот стиль состоит в следующем:

1. Устанавливается возможный сценарий для характерной роли черных дыр в реальной вселенной. [*Пример*: черная дыра, покоящаяся в межзвездной среде, будет аккрецировать межзвездный газ; падающий газ мог бы достаточно сильно нагреться, так, чтобы излучать значительно раньше того момента, когда он достигнет горизонта, и излучение могло бы быть настолько сильным, чтобы наблюдаться на Земле. Этот сценарий впервые был установлен Солпитером [174] и Зельдовичем [228] и детально изучен Шварцманом [185], Мишелем [131], Шапиро [179, 189], Принглом, Рисом и Пахольчиком [163], Межаросом [129] и Бисноватым-Коганом [20].]

2. Угадывается макроскопическое описание физики сценария и решаются соответствующие макроскопические уравнения, чтобы получить первый грубый вариант модели. [*Пример*: предполагается, что аккрецирующий газ может рассматриваться как идеальная гидродинамическая жидкость с плотностью $\rho(r)$, давлением $p(r)$ и радиальной скоростью $v(r)$, движущаяся в сферически-симметричном ньютоновском потенциале $\Phi=-GM/r$ черной дыры с массой M. Предполагают и надеются, что микроскопическая физика будет в конечном счете давать «политропическую» зависимость давления от плотности $p=K\rho^{\gamma}$ для некоторых констант K и γ. Это позволяет

[1]) Можно рекомендовать также учебники [234 — 238], журнал «Природа» и журнал «Успехи физических наук» — в той же последовательности.— *Прим. перев.*

отделить гидродинамические уравнения от уравнений для радиационного охлаждения и теплопереноса и на время отложить рассмотрение процессов охлаждения и теплопереноса. Затем решаются стандартные гидродинамические уравнения с граничными условиями на бесконечности, соответствующими «типичному» межзвездному газу — ионизованной, в значительной степени водородной плазме с температурой $T_\infty \sim 10^4$K и с

$$v_\infty = 0, \ \rho_\infty \sim 10^{-24} \ \text{г}\cdot\text{см}^{-3}, \ (p_\infty/\rho_\infty)^{1/2} \sim 10 \ \text{км}\cdot\text{с}^{-1}. \qquad (1)$$

Оказывается, что дыра слабо влияет на радиусах $r \gg R_A \equiv$ («радиус аккреции»), где R_A определяется соотношением

$$GM/R_A = p_\infty/\rho_\infty \approx (\text{скорость звука при } r = \infty)^2, \qquad (2)$$
$$R_A \approx 10^{14} \ \text{см} \ (M/M_\odot),$$

но что вблизи R_A она втягивает газ и при $r \ll R_A$ газ свободно падает, причем

$$\frac{1}{2} v^2 \approx GM/r, \quad \rho \approx \rho_\infty (R_A/r)^{3/2} \ \text{вообще}, \qquad (3)$$
$$T \approx T_\infty (R_A/r)^{3(\gamma-1)/2}, \quad \text{пока } p \sim \rho T.$$

(Здесь и далее M — масса центрального тела, измеряемая гравитационно, а $M_\odot = 2 \cdot 10^{33}$ г$=1{,}5$ км — масса Солнца.) Полная скорость аккреции массы есть

$$\dot{M} \approx 4\pi R_A^2 \rho_\infty (p_\infty/\rho_\infty)^{1/2} \sim (10^{11} \ \text{г}\cdot\text{с}^{-1}) (M/M_\odot)^2. \qquad (4)$$

Детали этого процесса впервые были рассмотрены в работах Хойла и Литтлтона [84] и Бонди [30], обзор дан в § 4.2 обзора Новикова и Торна [135].]

3. Следующий этап состоит в исследовании микроскопической физики сценария, чтобы выяснить, соответствует ли она макроскопической модели. Если нет, то выбирают новую модификацию макроскопической модели и повторяют описанный выше процесс. Если имеется грубое соответствие, то получают более детальные макроскопические уравнения, которые являются лучшим приближением микроскопической физики и решение которых может дать улучшенную модель. [*Пример*: спрашивается, справедливо ли для представленного выше решения гидродинамическое приближение, по крайней мере при $r \sim R_A$, где определяется скорость аккреции массы. Для проверки гидродинамического приближения радиус аккреции R_A сравнивается с расстоянием λ_p, которое должен пройти протон в межзвездной среде, прежде чем кулоновское рассеяние существенно отклонит его от прямолинейного движения:

$$\lambda_p \sim (Z \cdot 10^{12} \ \text{см}) (\rho/10^{-24} \ \text{г}\cdot\text{см}^{-3})^{-1} (T/10^4 \ \text{К})^2. \qquad (5)$$

Оказывается, что

$$\lambda_p/R_A \sim 0{,}1 \ (M/M_\odot)^{-1}. \qquad (6)$$

Это один из тех редких случаев, когда безразмерное отношение оказывается близким к единице. Если бы выполнялось неравенство $\lambda_p/R_A \ll 1$, то гидродинамическое приближение было бы справедливо и можно было бы спокойно развивать модель дальше. Если оказывается, что $\lambda_p/R_A \gg 1$, надо начинать все сначала, с новой макроскопической модели, в которой протоны являются независимыми, невзаимодействующими частицами, несущими с собой электронные облака (см. [15] и ссылки там). Но в нашем случае ничто экстремальное не подходит. В отчаянии принимаемся искать в списке физических процессов что-нибудь помимо кулоновского рассеяния, что могло бы сэкономить время и обеспечить законность гидродинамического приближения. Две вещи приходят в голову: рассеяние протонов на плазмонах (т. е. на коллективных возбуждениях плазмы) и прикрепление протонов к межзвездному магнитому полю. Исследование безразмерных отношений характерных длин показывает, что протон-плазмонное рассеяние в лучшем случае существенно на пределе точности, а сцепление с магнитным полем чрезвычайно важно:

$$\frac{\text{Ларморовский радиус протона}}{\text{Радиус аккреции}} \approx \frac{(10^8 \text{ см})\,(B/10^{-6}G)^{-1}}{(10^{14} \text{ см})\,(M/M_\odot)} \ll 1. \qquad (7)$$

Если магнитное поле вблизи R_A является запутанным (т. е. неоднородным в масштабах $l_B \ll R_A$), то оно будет обеспечивать физическое сцепление между протонами, и гидродинамическое приближение будет законным. Тогда можно продолжать работать с гидродинамической моделью, дополнив ее соответствующим макроскопическим описанием магнитотурбулентной плазмы и соответствующими уравнениями испускания синхротронного и тормозного излучения; детали см. в работах Шварцмана [185] и Шапиро [179, 180]. С другой стороны, если магнитное поле вблизи R_A очень однородно (т. е. $l_B/R_A \gg 1$), то нужно опять начинать все сначала, с макроскопической модели, в которой аккрецирующие, невзаимодействующие протоны плавно увлекаются магнитным полем, давая в конечном счете конфигурацию поля типа «песочных часов» с плазмой, скользящей вдоль силовых линий по направлению к перемычке песочных часов. В области этой перемычки плазма образует диск и постепенно скользит по силовым линиям к дыре, сильно нагреваясь по мере скольжения. Детали в работах Бисноватого-Когана и Рузмайкина [22, 23] и Бисноватого-Когана [20].]

Приведенные выше примеры иллюстрируют следующие характерные черты работы с черными дырами в астрофизике:

1. Она включает итерации «туда и обратно» между уравнениями макроскопической модели и микроскопической физикой, лежащей в основе этих уравнений. Итерации продолжаются до получения самосогласованности.

2. На каждой итерационной стадии следует тщательно следить, чтобы не пропустить процессы, которые могли бы оказаться настолько важными, что привели бы к незаконности модели (в приве-

денном выше примере это прикрепление к однородному межзвездному магнитному полю).

3. Часто встречаются «точки ветвления», где модель распадается на две совершенно различные формы в зависимости от того, какие предположения делаются об окрестности черной дыры (однородное или запутанное магнитное поле в приведенном выше примере), причем обе «ветви» могли бы существовать в реальной вселенной. Это ведет к множеству возможных моделей, причем каждая соответствует различному внешнему окружению черной дыры и/или различной массе черной дыры.

Все сценарии, которые мы описываем в оставшейся части этой статьи, будут основаны на моделях, которые были построены итеративными макроскопически-микроскопическими рассмотрениями. Однако при описании сценариев мы ограничимся минимумом деталей. Детали сделали бы эту статью слишком длинной.

3. ОДИНОЧНЫЕ ЧЕРНЫЕ ДЫРЫ, ВОЗНИКАЮЩИЕ ПРИ КОЛЛАПСЕ НОРМАЛЬНЫХ ЗВЕЗД

3.1. СМЕРТЬ ЗВЕЗД И РОЖДЕНИЕ ДЫР

В нашей Галактике звезды непрерывно рождаются и непрерывно умирают. С уверенностью известно, что они рождаются в процессе гравитационного сжатия плотных межзвездных газовых облаков; что после начальной стадии сжатия, длящейся примерно $5\cdot10^7$ лет для массы $M\sim M_\odot$ и меньше 10^5 лет для $M>10M_\odot$, в их недрах загорается водород; что затем они существуют за счет ядерного горения в течение $t>10^{10}$ лет, если $M<M_\odot$, и $t<2\cdot10^7$ лет, если $M>10M_\odot$, и что затем они умирают. (Детали этой картины можно найти в § 6.5—6.7 книги Клейтона [47] и в обзоре Айбена [86].) [1])

Детали агонии звезд известны плохо. Однако типы возможных остатков умерших звезд твердо установлены теорией. Они включают расширяющиеся газовые облака, возникшие в результате разрушения всей звезды или ее части (примером может служить Крабовидная туманность), белые карлики (известны сотни примеров, см., например, [72, 206]), нейтронные звезды (хорошо известные как источники энергии и регуляторы пульсаров и рентгеновских компонент в некоторых двойных рентгеновских источниках, см., например, [103, 121]) и черные дыры. Белые карлики не могут иметь массу более чем $M^{wd}_{\text{макс}}\approx1{,}4M_\odot$, если они вращаются медленно [43], или $\sim3M_\odot$, если они вращаются быстро и дифференциально [58]. Нейтронные звезды не могут превышать предела $M^{n*}_{\text{макс}}$, который в зависимости от способа рассмотрения оценивается в пределах между 1,3 и 2,5 $M_\odot$ (см., например, [5]) и который, по-видимому, ограничен соотношением

$$M^{n*}_{\text{макс}}<(5M_\odot)\,(\rho_0/\rho_{\text{яд}})^{-1/2} \qquad (8)$$

[1]) См. также работу Зельдовича и Новикова [239].— *Прим. перев.*

даже при наличии вращения (см., например, [76]). Здесь $\rho_{яд}=2\cdot10^{14}$ г·см$^{-3}$— плотность материи внутри обычного атомного ядра, а $\rho_0\sim(0,5\div5)\rho_{яд}$ — плотность, при которой, как полагают, современная теория ядерной материи в больших масштабах может приводить к серьезным ошибкам. Измеренные массы белых карликов и нейтронных звезд согласуются с этими теоретическими ограничениями [73, 164, 216]. Черные дыры, являющиеся конечным продуктом звездной эволюции, могли бы иметь массы в диапазоне от $M^{n*}_{макс}$ вплоть до масс самых тяжелых нормальных звезд $M\sim100\ M_\odot$ (см., например, [47]), но они, вероятно, не могут иметь масс, меньших $\sim M^{n*}_{макс}$, так как ядерные силы препятствуют сжатию таких малых количеств материи до плотностей черной дыры.

Конечная судьба данной звезды зависит не только от ее массы при рождении, но и от того, как много массы она ухитрится потерять за свою жизнь. Тот факт, что должна происходить значительная потеря массы, следует из наличия белых карликов ($M<1,4\ M_\odot$) в Гиадах — звездном скоплении столь молодом, что в нем до сих пор умерли только звезды с начальной массой $M>2,1\ M_\odot$ [7]. Более того, утекающая масса («звездный ветер») наблюдается спектроскопически для большого количества звезд, причем для некоторых звезд главной последовательности (живущих за счет горения водорода) массы $M\gtrsim20M_\odot$ измеренный ветер достаточно силен, чтобы значительно уменьшить массу звезды, прежде чем в ее недрах исчерпается водородное горючее [106]. Некоторые наиболее аргументированные современные предположения относительно количества массы, теряемой различными звездами, изложены в обзоре Вайдемана [217]. Однако они представляют собой скорее догадки, возникающие на стадии изучения. Приемлемо, хотя и маловероятно, что все звезды *могли бы* сбрасывать свою массу до предела $M<M^{n*}_{макс}$ еще до начала процесса умирания, избегая тем самым судьбы черных дыр; и также приемлемо, хотя и маловероятно, что все звезды с массой, превышающей $M^{n*}_{макс}$ в конце жизни, могли бы выбросить достаточно массы на стадии агонии, чтобы стать нейтронными звездами. Ни имеющиеся наблюдательные данные, ни современная теория не могут исключить этих возможностей. Имеется печальный комментарий к теории, который состоит в том, что если игнорировать наблюдательные ограничения, то можно допустить для всех звезд даже уменьшение массы до значений, лежащих ниже предела белых карликов $M^{wd}_{макс}$, что означает невозможность образования нейтронных звезд. Действительно, вскоре после открытия пульсаров появилась теоретическая работа [4], в которой говорилось, что нейтронные звезды вообще никогда не могут образоваться в природе из-за того, что углеродно-детонированный ядерный взрыв должен разрушить любую звезду с массой, превышающей $M^{wd}_{макс}$, прежде чем она могла бы коллапсировать. Только недавно теория была адекватно приспособлена для рассмотрения процесса образования нейтронных звезд (см. [123] и ссылки там).

В другом предельном случае, который также согласуется с современной теорией и наблюдательными данными, вполне вероятно, что потери массы пренебрежимо малы для большинства звезд с массой, превышающей $M^{n*}_{\text{макс}} \sim 2{,}5\ M_{\odot}$. Если это так, то наблюдаемое в окрестности Солнца звездное население позволяет оценить скорость образования черных дыр в $\sim 1{,}5 \cdot 10^{-10}$ лет$^{-1}$ (пс^2 от галактической плоскости)$^{-1}$ [142]. Экстраполируя назад во времени на возраст Галактики 10^9 лет и не предполагая существенного увеличения скоростей черных дыр при рождении, обусловленного эжекцией момента [17, 18], мы приходим к выводу, что ближайшая к Земле черная дыра звездной массы находится сегодня на расстоянии ~ 5 пс. Экстраполяция на всю галактическую плоскость приводит к тому, что черные дыры должны рождаться в нашей Галактике со скоростью приблизительно одна дыра каждые пять лет и что полное число черных дыр $\sim 2 \cdot 10^9$. Для сравнения: число вспышек Сверхновых в нашей Галактике — из исторических наблюдений — колеблется между одной в шесть лет и одной в 25 лет, число вспышек Сверхновых во внешних галактиках нашего типа и массы — между одной в 10 лет и одной в 40 лет [197]; скорость рождения пульсаров в нашей Галактике оценивается из наблюдательных данных от одного пульсара в 8 лет до одного пульсара в 40 лет [198].

Приведенные выше оценки числа черных дыр не следует принимать слишком всерьез. Наряду с существованием неопределенностей, обусловленных потерями массы, весьма рискованной является также экстраполяция назад во времени или на другие области Галактики. Звездные популяции, существовавшие раньше или существующие сегодня в центральных и внешних областях нашей Галактики, могут заметно отличаться от наблюдаемого сегодня вблизи Земли звездного населения (см., например, [204, 210]).

Вернемся к деталям процесса рождения черных дыр. Черные дыры не могут быть образованы в результате квазистационарного сжатия звезд, так как гравитационное ускорение в стационарной звезде на стадии образования горизонта должно быть бесконечным. Поэтому дыры наверняка должны образовываться при динамическом звездном коллапсе. Однако это почти единственная вещь, известная наверное.

Для идеализированного случая сферического коллапса имеется много численных моделей процесса образования черной дыры, в которых используются реалистические уравнения состояния (см., например, [122]). Однако в сферически-симметричной картине существуют значительные неопределенности, связанные с нерешенными вопросами переноса тепла нейтрино (см., например, [213]), а когда начинают рассматривать еще магнитные поля, вращение и другие отклонения от сферичности — неизвестно почти ничего. Численные расчеты несферического коллапса проводятся сейчас несколькими исследовательскими группами (в Ливерморе, Гарварде, Чикаго и Техасе), но в ближайшем будущем появление достоверной картины

маловероятно. Среди важных вопросов, остающихся открытыми, назовем следующие: могут ли черные дыры образовываться в ядрах каких-либо Сверхновых? Какое количество массы уносится из внешних слоев коллапсирующей звезды? Сколько испускается гравитационного излучения? Как зависят ответы на эти вопросы от массы и углового момента предсверхновой звезды? Какова роль магнитных полей [24, 124]?

Может быть, наибольшие надежды можно возлагать на детектирование и изучение гравитационного излучения, возникающего в процессе образования черной дыры. Форма гравитационного импульса даст непосредственную информацию о динамическом поведении звездного ядра до и в процессе образования горизонта, а также о массе и угловом моменте новорожденной дыры. (Масса и угловой момент определяют комплексные собственные частоты нормальных мод пульсаций черной дыры и тем самым должны определять поведение гравитационных импульсов (их форму) в более поздние моменты времени). Оценка напряженности волн является весьма неопределенной, однако в случае сильно несферического коллапса представляется вероятным, что они будут нести энергию, равную нескольким процентам от массы дыры [56, 57] в широкополосном импульсе с максимумом на частоте

$$\nu \sim (10 \text{ кГц})(M/M_{\odot}).$$

Такие импульсы от источника, расположенного в нашей собственной Галактике (расстояния $\sim$10 кпс), могли бы без труда детектироваться самой чувствительной гравитационно-волновой антенной, действующей с 1976 г. Конечной целью следующих поколений детекторов является наблюдение и изучение импульсов, приходящих с расстояния $\sim$10 Мпс, соответствующего скоплению галактик M101, где имеется примерно одна Сверхновая в год; при увеличении этого расстояния число Сверхновых возрастает пропорционально кубу расстояния. Дальнейшие подробности можно найти в обзоре Торна [200].

3.2. АККРЕЦИЯ МЕЖЗВЕЗДНОГО ГАЗА НА ЧЕРНУЮ ДЫРУ

Если черная дыра образовалась в результате звездного коллапса в нашей Галактике, то у нее почти нет шансов когда-нибудь (в течение $\sim 10^{10}$ лет) столкнуться с какой-нибудь другой дырой или звездой. Исключение, возможно, составляют только черные дыры, образующиеся в двойных системах (разд. 4), в шаровых скоплениях (разд. 5) или во внутренних областях ($\sim$несколько парсек) галактического ядра (разд. 6). С другой стороны, у одиночной черной дыры в межзвездном пространстве ничтожно мало шансов проявить себя в качестве гравитационной линзы по отношению к излучению более удаленной звезды [169]. Поэтому, вероятно, единственная на-

дежда обнаружить ее — наблюдение электромагнитного излучения межзвездного газа, падающего хаотически к горизонту черной дыры.

Модели аккреции межзвездного газа на изолированную черную дыру за последние десять лет развиты детально. Первые шаги на пути построения таких моделей были описаны в разд. 2 [уравнения (1) — (7)]. Как мы видели, решающую роль здесь играет межзвездное магнитное поле. Оказывается, что независимо от упорядоченности или запутанности поля в масштабе радиуса аккреции R_A поле обеспечивает связывание протонов, достаточное для того, чтобы сделать течение гидродинамическим при R_A. В результате скорость аккреции массы всегда определяется формулой (4), если дыра покоится в межзвездной среде, или простым обобщением этой формулы на случай движущейся дыры:

$$\begin{aligned} R_A &\approx 10^{14}\ \text{см}\,(M/M_\odot)\,(T_\infty/10^4\ \text{К})^{-1}(1+\mu^2)^{-1}, \\ \dot M &\approx 4\pi R_A^2\rho_\infty\,(p_\infty/\rho_\infty)^{1/2}(1+\mu^2)^{-1/2} \approx 10^{11}\ \text{г}\cdot\text{с}^{-1}\,(M/M_\odot)^2\,\xi, \qquad (9) \\ \xi &\equiv (\rho_\infty/10^{-24}\ \text{г}\cdot\text{см}^{-3})\,(T_\infty/10^4\ \text{К})^{-3/2}(1+\mu^2)^{-3/2}. \end{aligned}$$

Здесь μ — число Маха для движения черной дыры сквозь межзвездное пространство

$$\mu = \frac{(\text{скорость дыры})}{(p_\infty/\rho_\infty)^{1/2}} = \frac{(\text{скорость дыры})}{10\ \text{км}\cdot\text{с}^{-1}}\left(\frac{10^4\ \text{К}}{T_\infty}\right)^{1/2}. \qquad (10)$$

Отметим, что при больших значениях чисел Маха, т. е. $\mu \gg 1$, формула (9) получается из формулы для случая «покоя дыры» простой заменой скорости звука $(p_\infty/\rho_\infty)^{1/2}$ на скорость дыры. Фактор $(1+\mu^2)$ выбирается так, чтобы иметь простую интерполяцию между случаями малой и большой скорости [30].

Ниже радиуса аккреции течение газа приближается к свободному падению со скоростью $v \sim r^{-1/2}$, плотностью $\rho \sim r^{-3/2}$, температурой $T \sim r^{-1}$ и (как следствие сохранения потока) магнитным полем $B \sim r^{-2}$; см. (3). Вне и вблизи R_A магнитным давлением можно было пренебречь по сравнению с тепловым давлением; магнитное поле служило лишь для связывания протонов друг с другом с помощью ларморовского вращения. Однако ниже R_A

$$\begin{aligned} B^2/8\pi \sim r^{-4}, \quad p \sim \rho T \sim r^{-5/2}, \\ (GM/r)\,\rho \sim r^{-5/2}. \qquad (11) \end{aligned}$$

Это означает, что магнитное давление очень быстро подавляет тепловое давление, и вскоре после этого плотность магнитной энергии становится сравнимой с плотностью гравитационной энергии. Впоследствии характер течения будет определяться конкуренцией магнитных и гравитационных сил и будет сильно отличаться от простого радиального свободного падения.

Разнообразие различных моделей, построенных для такого течения, соответствует разнообразию различных условий на бесконечности. Если межзвездное поле однородно в масштабах $l_B \gg R_A$

(наиболее вероятный случай) и если в сдерживаемом магнитными силами потоке не развивается неустойчивостей, мы получаем «модель песочных часов», кратко описанную в разд. 2 и развитую детально Бисноватым-Коганом и Рузмайкиным [22, 23]; см. также [20, 145, 156, 194].

В случае запутанных магнитных силовых линий на бесконечности или при наличии неустойчивостей, приводящих к их запутыванию, течение газа становится магнитотурбулентным. Тогда существующие модели без достаточного обоснования предполагают следующее:

1. Магнитные поля существенно замедляют падение, так что примерно половина высвобождаемой гравитационной энергии переходит в энергию падения и примерно половина — в магнитную энергию и энергию турбулентного движения:

$$\frac{1}{2}\rho v^2 \sim \frac{B^2}{8\pi} + \frac{1}{2}\rho v^2_{\text{турб}} \sim \frac{1}{2}\rho\frac{GM}{r}; \qquad 4\pi r^2 \rho v = \dot{M} = \text{const}. \tag{12}$$

2. Перезамыкание силовых линий спасает магнитное поле от возрастания с соответствующей сохранению потока скоростью $B\sim r^{-2}$ и дает ему возможность возрастать медленнее, со скоростью $B\sim r^{-5/4}$, определяемой соотношением (12).

3. Потери магнитной энергии при перезамыкании силовых линий — и потери энергии турбулентного движения за счет турбулентной вязкости — переходят в тепло (т. е. в случайную кинетическую энергию протонов и электронов). Результирующая скорость нагревания на единицу объема определяется как

$$\Gamma \sim \frac{1}{2}\left(\frac{v}{r}\right)\left(\frac{B^2}{8\pi} + \frac{1}{2}\rho v^2_{\text{турб}}\right) \sim \left(\frac{GM}{r^3}\right)^{1/2}\rho\left(\frac{GM}{r}\right). \tag{13}$$

Эти предположения кажутся разумными, однако теория сильной магнитной турбулентности находится в столь примитивном состоянии, что строго оправдать их невозможно.

После того как были сделаны эти предположения, магнитотурбулентные модели расщепились на несколько различных типов в зависимости от 1) количества углового момента в аккрецируемом газе (достаточно или слишком мало, чтобы обеспечить центробежное замедление радиального падения и тем самым образование аккреционного диска); 2) степени остывания за счет циклотронного и синхротронного излучений (достаточно или слишком мало, чтобы конкурировать с магнитотурбулентным нагреванием); 3) оптической толщины падающего газа для синхротронного и циклотронного излучений. Каждый тип модели зависит от условий в аккрецирующем газе на бесконечности (ρ_∞, T_∞, B_∞ и наличие вихрей). Модели с центробежным замедлением (большая «завихренность» на бесконечности) обсуждались Солпитером [174] и Шварцманом [185]. Раз-

личные случаи без центробежного замедления исследованы Шварцманом [185], Шапиро [179, 180] и Межаросом [129].

Полная светимость соответствующей модели может быть выражена в терминах универсальной скорости аккреции (9) и эффективности модели ε, характеризующей переход высвобождаемой гравитационной энергии, $\dot{E}_{грав}=\dot{M}c^2$, в уходящее излучение

$$L=\varepsilon\dot{M}c^2=(10^{-2}L_\odot)(\varepsilon/0{,}5)(M/M_\odot)^2\xi. \quad (14)$$

Для модели «песочных часов» и для модели с центробежным замедлением эффективность ε высока: $\varepsilon\sim$ (от 0,05 до 0,5). Для других моделей она ниже, например Межарос [129] дает оценку $\varepsilon\sim10^{-4}$ и $L\sim10^{-4}L_\odot$ для случая $M=10M_\odot$, $\rho_\infty=10^{-24}$ г·см$^{-3}$, $T_\infty=10^4$ К, $\xi=1$ и магнитотурбулентного течения без вращательного замедления.

Во всех моделях уходящее излучение складывается в основном из циклотронного и синхротронного излучений (электронов, вращающихся в магнитном поле). В большинстве моделей основное количество излучения испускается из области вблизи шварцшильдовского радиуса. Если это излучение приблизительно чернотельное, то оно должно иметь пик на частоте $\nu_{макс}$, определяемой из соотношений

$$h\nu_{макс}\approx kT_{макс},\quad 4\pi(3GM/c^2)^2\sigma T^4_{макс}\approx L, \quad (15)$$

т. е.

$$h\nu_{макс}\approx 60\ эВ\,(\varepsilon\xi)^{1/4}. \quad (16)$$

В ультрафиолетовой части спектра этому соответствует $\varepsilon\xi\sim1$, в видимой части спектра —$\varepsilon\xi\sim10^{-4}$. Если спектр отличен от спектра чернотельного излучения, он будет иметь максимум в области частот, равных или бóльших чем 60 эВ $(\varepsilon\xi)^{1/4}$.

Большинство детально разработанных моделей дают спектры с пиком вблизи частоты 60 эВ$(\varepsilon\xi)^{1/4}$, экспоненциальным падением выше этой частоты и медленным падением ниже (типичный результат $dL/d\nu\propto\nu^{от\ 0\ до\ 0,3}$. Поэтому черная дыра с массой в несколько масс Солнца, аккрецирующая межзвездный газ, будет обнаруживать себя на оптических частотах как «звезда» низкой, но не пренебрежимо малой светимости с довольно плоским, невыразительным спектром. Шварцман [185], создатель первой модели, который пришел к этому выводу, отметил сходство такого объекта с *DC*-белыми карликами (т. е. белыми карликами без линий в спектрах) и предположил, что некоторые из наблюдаемых *DC*-белых карликов могли бы оказаться на самом деле черными дырами. Он также обратил внимание на характерный признак, которым могло бы обладать излучение, идущее из окрестности черной дыры: оно могло бы флуктуировать с характерным временем

$$\tau\sim(\text{несколько единиц})\times\left(\frac{GM}{c^3}\right)\sim(\text{несколько единиц})\times$$
$$\times10^{-5}\ \text{с}\left(\frac{M}{M_\odot}\right). \quad (17)$$

К сожалению, детектирование таких флуктуаций в объекте столь низкой светимости на разумных расстояниях ($\gtrsim$100 пс) требует очень большого телескопа, а предсказания так неопределенны, что никто пока не захотел предоставить такой телескоп на достаточно большое время для поиска черных дыр. Тем не менее вполне возможно, что будущие исследования смогут отождествить такие объекты.

Кроме аккреции «стандартного» межзвездного газа можно представить себе другие условия аккреции на изолированную черную дыру. В одном предельном случае это была бы аккреция в галактическом гало, где межзвездная плотность гораздо ниже чем 10^{-24} г$\times$ $\times$см$^{-3}$ и где возникающее излучение почти несомненно не наблюдаемо, пока масса дыры не станет равна $M \gg 100\, M_{\odot}$. В другом предельном случае возможна аккреция из очень плотного облака газа, т. е. газа, выброшенного в течение первоначального процесса образования дыры; в этом случае мог бы возникнуть рентгеновский источник высокой светимости ($L \sim 10^6 L_{\odot}$) [130]. Самый экстремальный случай — это черная дыра, которая образуется в ядре массивной звезды без сбрасывания диффузной звездной оболочки. (Согласно «житейской мудрости», такое никогда не случится, но наверняка этого случая исключить нельзя.)

Для случая дыры внутри звезды по крайней мере три исследовательские группы (в Калифорнийском технологическом институте, Кембридже и Мюнхене) пытались построить стационарные модели аккреции, но все оказалось безуспешным. Кажется правдоподобным, что аккреция могла бы происходить в стремительном экспоненциальном режиме с дырой, проглатывающей всю звезду за время свободного падения. И столь же правдоподобно, что дыра «съест» звезду только за время $\sim 10^7$ или 10^8 лет, что аккреция будет происходить в колебательном по времени режиме и что продуктом такой аккреции будет усредненная по времени светимость $L_{\text{эдд}} \approx$ $\approx (4 \cdot 10^4 L_{\odot}) \times (M/M_{\odot})$. Эта задача требует большой работы. Детали первоначальных неудач и частичных успехов см. в работах [16] и [96].

4. ЧЕРНЫЕ ДЫРЫ В ДВОЙНЫХ СИСТЕМАХ

4.1. ВВЕДЕНИЕ

Примерно 50% всех звезд рождается в двойных системах: это справедливо и для больших звезд, и для маленьких. И в 50% всех двойных систем звезды расположены достаточно близко, чтобы существенно взаимодействовать в процессе эволюции. Интересная для нас возможность состоит в том, что более массивная звезда в двойной системе будет быстро расходовать свое ядерное горючее и коллапсировать, образуя черную дыру, и что затем ее менее массивный компаньон будет поставлять достаточно газа на дыру, обеспечивая достаточно большую светимость.

Грубая оценка максимально возможной светимости — это «эддингтоновский предел» — значение L, для которого действие гравитации на аккрецируемый газ точно сбалансировано направленным наружу давлением фотонов, которые рассеивают электроны газа:

$$\frac{GM\rho}{r^2} = \frac{L_{эдд}/c}{4\pi r^2} \sigma_T \frac{\rho}{m_H}, \tag{18}$$

т. е.

$$L_{эдд} = \frac{4\pi GcMm_H}{\sigma_T} = (1\cdot 10^{38} \text{ эрг}\cdot\text{с}^{-1})\left(\frac{M}{M_\odot}\right). \tag{19}$$

(Здесь σ_T — томсоновское сечение рассеяния фотонов электронами, m_H — масса атома водорода и ρ/m_H — численная плотность электронов в аккрецирующем газе, который предполагается состоящим в основном из ионизованного водорода.) Если основная часть светимости L приходит из области вблизи шварцшильдовского радиуса, то типичные фотоны должны иметь энергию, бо́льшую или порядка энергии, соответствующей чернотельному излучению:

$$h\nu \gtrsim kT_{BB}, \quad 4\pi (3GM/c^2)^2 \sigma T_{BB}^4 \approx L, \tag{20}$$

т. е.

$$h\nu \gtrsim 3\text{кэВ}\,(M/M_\odot)^{-1/2} (L/L_{эдд})^{1/4}. \tag{21}$$

Таким образом, черная дыра, аккрецирующая газ в тесной двойной системе, является многообещающим источником рентгеновского излучения высокой светимости. То же самое верно для нейтронной звезды и в меньшей степени для белого карлика.

Идея поисков рентгеновского излучения от черных дыр и нейтронных звезд в двойных системах пришла в голову многим астрофизикам одновременно в 1966 г. (см. историю вопроса в работе [32] и ранние публикации [136, 184]). Эта идея оправдала себя в 1971 г., когда исследовательская группа Джиаккони запустила первый рентгеновский спутник Земли «Ухуру» и открыла рентгеновское излучение пульсарного типа со светимостью $L \sim L_{эдд}$, приходящее из нескольких двойных систем (см. обзор [66]). Пульсирующие рентгеновские потоки очень быстро и убедительно были приписаны нейтронным звездам [54, 68, 105, 162]; они не могли бы приходить от черных дыр, так как теория требует, чтобы дыры были аксиально-симметричными, и тем самым не позволяет их вращению действовать в качестве часового механизма, регулирующего импульсы, точно следующие во времени.

С другой стороны, если нейтронные звезды могут оказаться в двойных системах и могут испускать рентгеновское излучение при аккреции, то нет никаких видимых причин, по которым то же самое не могли бы делать черные дыры. Поэтому заманчиво было предположить, что некоторые из Ухуровских непульсирующих источников могут быть черными дырами. Наиболее убедительный способ отличить черные дыры от нейтронных звезд и белых карликов — это взвесить их; что-либо более тяжелое, чем $M_{макс}^{n*}$, должно быть черной дырой. Взвешивание было выполнено, хотя и не очень точно,—

и выдвинуло многообещающего кандидата в черные дыры: рентгеновский источник Cygnus X-1. В разд. 4.5 процедура взвешивания описывается детально наряду с другими свидетельствами в пользу предположения (но еще не доказательства!), что X-1 и Circinus X-1 — черные дыры.

Прежде чем обсуждать наблюдательную ситуацию, представим необходимые для этого теоретические основания, описывая в разд. 4. 2 эволюцию массивных двойных систем и в разд. 4.3 теорию сбрасывания газа в дыру ее компаньоном и наконец в разд. 4.4— модели течения газа вблизи дыры и генерации рентгеновского излучения.

4.2. ЭВОЛЮЦИЯ ТЕСНЫХ ДВОЙНЫХ СИСТЕМ

Чем массивнее звезда, тем быстрее она будет сжигать свое ядерное горючее и тем скорее она умрет. Этот факт наводит на мысль, что в двойных системах мертвый, аккрецирующий и испускающий рентгеновское излучение объект должен быть более массивным, чем его живой, поставляющий массу компаньон. Однако верно как раз обратное. Во всех измеренных случаях мертвый объект менее массивен.

Простое объяснение этого факта следует из теории эволюции тесных двойных систем. Рассмотрим две молодые звезды на орбите относительно друг друга с массами M_1 и M_2 и радиусами R_1 и R_2, на расстоянии $D > R_1 + R_2$. Более массивная звезда 1 (M_1, R_1) называется главной звездой; менее массивная — «второстепенной». Когда главная звезда израсходует водород в своем центре, она переходит к сжиганию водорода в оболочках; ее центр несколько сжимается, а поверхность расширяется. По мере того как оболочка, в которой происходит горение водорода, постепенно движется наружу, ядро звезды продолжает сжиматься, а ее наружные слои продолжают расширяться, постепенно превращая звезду в красный гигант. Наконец достигается критический радиус этого расширения

$$\begin{aligned} R_{\text{крит}} &\approx [0{,}38 + 0{,}2 \lg_{10} (M_1/M_2)]\, D, \quad \text{если } 0{,}3 < M_1/M_2 < 20, \\ &\approx 0{,}462\, (1 + M_2/M_1)^{-1/3} D, \quad \text{если } M_1/M_2 < 0{,}8, \end{aligned} \qquad (22)$$

при котором главная звезда вторгается в источник гравитационного потенциала своего компаньона. Когда R_1 достигает $R_{\text{крит}}$, говорят, что главная звезда «заполняет свою полость Роша» и начинается перетекание, которое происходит столь интенсивно, что к моменту его прекращения главная звезда оказывается менее массивной, чем второстепенная. Если главная звезда в результате не уменьшает свою массу ниже чандрасекаровского предела $M^{wd}_{\text{макс}}$, она будет коллапсировать, образуя нейтронную звезду или черную дыру, прежде чем вторая звезда сильно эволюционирует, и если сбрасывание массы при коллапсе не разрушит двойную систему, результатом бу-

дет нейтронная звезда или черная дыра на орбите вокруг более массивной, нормальной звезды.

Детальные модели делают такой сценарий более правдоподобным (см., например, [212]). Однако в этих моделях имеются некоторые трудности. В моделях, построенных для объяснения образования наблюдаемых рентгеновских двойных, начальное разделение D не должно быть гораздо больше, чем R_1+R_2. В результате, когда главная звезда расширяется, она сбрасывает газ во вторичную звезду быстрее, чем вторичная звезда может его принимать; вторичная звезда расширяется, приходя в контакт с главной звездой; в этой контактной системе масса продолжает перетекать из главной звезды на вторичную, но вскоре после установления контакта кончается сфера применимости вычислительных методов, используемых авторами модели. Конечный результат предсказать невозможно; о нем можно лишь делать заключение, исходя из наблюдаемого существования рентгеновских двойных. О деталях трудностей таких моделей и о некоторых соображениях относительно эволюции контактной системы см. в [62, 101]; обзор теории эволюции двойных систем см. в работе [199] и указанных там ссылках.

Насколько массивной должна быть главная звезда, чтобы превратиться в черную дыру? Ответ будет зависеть от оценки переноса массы главной звездой на ее стадии гиганта, которая в свою очередь будет зависеть от начального отношения масс M_1/M_2 и расстояния D. В настоящее время эта зависимость неизвестна в силу неопределенностей, касающихся эпохи контакта. Неизвестна также вероятность того, что двойная система выдержит возмущения, возникающие, когда в процессе рождения дыры выбрасывается масса и испускается гравитационное излучение [70]. Поэтому можно думать, что четвертая часть всех черных дыр Галактики была рождена в тесных двойных системах, избежавших разрушения в процессе рождения черных дыр, и точно так же можно думать, что черные дыры никогда не встречаются в тесных двойных системах.

4.3. СБРАСЫВАНИЕ ГАЗА НА ДЫРУ ЗВЕЗДОЙ-КОМПАНЬОНОМ

Рассмотрим черную дыру, которой удалось образоваться в тесной двойной системе. До тех пор пока вторая звезда сохраняет возраст звезды «главной последовательности» (т. е. пока она находится на ранних стадиях горения водорода в центре), она будет поставлять на дыру очень мало массы. Однако по мере старения второй звезды происходят два явления: 1) она расширяется в размерах до тех пор, пока не переполнит свою полость Роша и не начнет сбрасывание сильного потока газа на дыру; 2) она становится все более яркой, и ее возрастающая яркость ведет к возрастающе сильному звездному ветру горячей плазмы с ее поверхности. Дыра может захватывать часть этого ветра. Когда дыра аккрецирует часть ветра или потока из полости Роша, может случиться третье явление;

3) сильный рентгеновский поток, ударяясь о поверхность звезды, может существенно видоизменить поток массы от звезды, тем самым видоизменяя рентгеновский поток и, возможно, даже образуя саморегулируемую систему. Конечно, все эти три процесса могут происходить и в случае нейтронных звезд в роли главной звезды так же, как и в случае черной дыры.

Когда были открыты рентгеновские двойные, среди теоретиков бушевали страсти вокруг относительной важности переполнения полости Роша, звездного ветра и саморегулирования; см., например, [54, 120, 158, 176, 211] и цитируемые там работы. Хотя эти дебаты еще не закончены, по-видимому, достигнуто предварительное согласие в нескольких пунктах.

1. Саморегуляция, вероятно, не является доминирующим фактором вблизи вторичной звезды, хотя она может играть важную роль вблизи черной дыры или нейтронной звезды [120].

2. Минимальная скорость аккреции

$$\dot{M}_{\text{мин}} \sim 3\,(L/c^2) \sim 3\cdot 10^{-10} M_{\odot}\,\text{год}^{-1} \tag{23}$$

требуется для генерации наблюдаемого рентгеновского излучения со светимостями $L \gtrsim 3000 L_{\odot}$ [211].

3. Аккрецирующая нейтронная звезда не сможет принять аккрецируемую массу, если скорость аккреции превышает величину

$$\dot{M}_{\text{макс}} \sim 10 L_{\text{эдд}}/c^2 \sim 10^{-8} M_{\odot}\,\text{год}^{-1}\,(M/M_{\odot}). \tag{24}$$

Значительно большие скорости переноса массы, вероятно, будут гасить рентгеновский источник; см. ниже, а также [211].

4. Звездный ветер, достаточно сильный для производства $\dot{M}_{\text{мин}}$, встречается только в звездах главной последовательности с массой, превышающей $45 M_{\odot}$ (звезды спектрального класса *Of*), и в звездах с массами $M \gtrsim 20 M_{\odot}$, исчерпáвших свои запасы водорода в центре и теперь расширяющихся на стадии гиганта (звезды, которые классифицируются спектроскопически как «сверхгиганты более ранние, чем тип B1») [211].

5. Звездного ветра хватает только на (несколько единиц)$\cdot 10^5$ лет в редком случае массивной звезды главной последовательности и только на (несколько единиц)$\cdot 10^4$ лет в более обычном случае «гиганта» [211].

6. Только на самых ранних стадиях переполнения полости Роша величину $\dot{M}$ можно удержать ниже $\dot{M}_{\text{макс}}$; раз достигнуто перетекание, развивается неустойчивость и уносит массу вторичной звезды со скоростью, заключенной между $\sim 10^{-6} M_{\odot}\,\text{год}^{-1}$ (для $M_2 \approx 2 M_{\odot}$) и $\sim 10^{-3} M_{\odot}\,\text{год}^{-1}$ (для $M_2 \approx 20 M_{\odot}$) [211].

7. Ранняя стадия перетекания с малым $\dot{M}$ будет продолжаться, если вращение вторичной звезды медленнее, чем ее орбитальное движение [176].

8. Стадия такого перетекания длиннее в наиболее тесных двойных, где перетекание начинается до того, как центральный водород полностью исчерпается («случай А»), чем в менее тесных двойных, где оно начинается на стадии горения водорода в оболочках («случай В») [176].

9. В более благоприятном случае А полная продолжительность перетекания в условиях $\dot{M}_{мин} \leqslant \dot{M} \leqslant \dot{M}_{макс}$ составляет величину $\sim 10^6$ лет для $M_2 \approx 2M_\odot$ и $\sim 10^4$ лет для $M_2 \approx 20M_\odot$ [176].

Как обсуждалось в 4.2, вторичная звезда обычно гораздо массивнее, чем сама дыра, т. е. $M_2 \gtrsim 10M_\odot$. Если это так, то перечисленные выше пункты подразумевают время жизни двойной системы в стадии рентгеновской двойной от $\sim 10^4$ до 10^5 лет по сравнению с временем жизни вторичной звезды на главной последовательности $\sim 10^7$ лет и современным возрастом галактики $\sim 10^{10}$ лет. Это обстоятельство вместе с существованием двух хороших кандидатов в черные дыры среди наблюдаемых рентгеновских двойных (разд. 4.5) означает, что наша Галактика может содержать $\gtrsim 300$ черных дыр с молодыми тесными компаньонами главной последовательности и $\gtrsim 3 \cdot 10^6$ мертвых двойных с черными дырами. Где они в небе?

Дыры с компаньонами главной последовательности очень трудно идентифицировать наблюдательно (см. разд. 4.5). Мертвые рентгеновские двойные, вероятно, отождествить еще труднее, хотя на самом деле и неизвестно, что искать. Смерть двойной системы с черной дырой может начаться, когда дыра начинает прятаться за неудержимо растущим переполнением полости Роша ее компаньона. Дыра может тогда найти свою орбиту под увеличенной поверхностью своего компаньона. Вообще говоря, не ясно, будет ли дыра затем двигаться по спирали к центру своего компаньона, образуя звезду с чернодырным ядром, которую теоретики не могут моделировать (разд. 3.2), или дыра будет способствовать сбрасыванию оболочки компаньона, но позволит ядру бодро эволюционировать к вымиранию в белого карлика, нейтронную звезду или черную дыру. В первом случае конечным продуктом была бы одна большая черная дыра. Во втором — это была бы тесная двойная система из двух компактных объектов (подобная «двойному пульсару» [85]), которая в конце концов слилась бы в один компактный объект под действием реакции гравитационного излучения. В любом случае смертные муки двойной системы могли бы оказаться интересными в наблюдательном отношении — если бы можно было угадать, что искать. Предварительные усилия по анализу этих результатов в случае двойной системы с нейтронной звездой сделаны Таамом, Боденхаймером и Острайкером [196], а также Торном и Житков [201]. Моделям с черной дырой до сих пор было уделено мало внимания.

4.4. МОДЕЛИ ТЕЧЕНИЯ ГАЗА НА ДЫРУ И ГЕНЕРАЦИИ РЕНТГЕНОВСКИХ ЛУЧЕЙ

С 1971 г., когда астрономы осознали, что Cygnus X-1 мог бы быть черной дырой, астрофизики посвятили много усилий моделированию течения газа на дыру и расчету свойств рентгеновских лучей, которые при этом испускаются. Полуколичественное согласие с наблюдениями достигнуто. Однако нет гарантии, что модели близко походят на реальность главным образом потому, что могут быть построены весьма различные модели, дающие одинаково хорошее согласие с наблюдениями.

Имеются две главные категории моделей: модели радиального падения и модели аккреционного диска. Какие из них правильны, зависит от того, насколько большим угловым моментом обладает аккрецирующий газ. Когда угловой момент j на единицу массы значительно превосходит величину

$$j_{\text{крит}} = (2GM_1/c^2)\,c, \tag{25}$$

центробежные силы будут определять характер течения и образовывать аккреционный диск на радиусах, больших по сравнению с горизонтом (M_1 — масса дыры, отмеченная индексом «1», так как она была главной звездой системы). Когда $j \lesssim j_{\text{крит}}$, центробежные силы будут оказывать малое влияние, и газ будет падать на дыру квазирадиально. Величина углового момента зависит от деталей выброса массы второй звездой. В случае переполнения полости Роша газ имеет почти такой же удельный угловой момент относительно центра дыры, как сама вторая звезда, поэтому $j \gg j_{\text{крит}}$, и аккреционный диск образуется на очень больших радиусах. В случае аккреции из ветра скорости v_w дыра при сверхзвуковом движении сквозь ветер создает своим гравитационным притяжением ударный фронт радиусом

$$R_S \approx 2GM_1/v_w^2. \tag{26}$$

Вещество, пересекающее этот ударный фронт, падает на дыру с чистым удельным угловым моментом

$$j \approx \left(\frac{R_A}{D}\right) R_A v_{\text{орб}}, \tag{27}$$

где D — расстояние между дырой и центром второй звезды и $v_{\text{орб}} \approx \approx (GM_2/D)^{1/2}$ — орбитальная скорость дыры относительно второй звезды. Для разумных значений параметров (например, параметров, соответствующих Cygnus X-1, где $M_1 \approx 10 M_\odot$, $M_2 \approx 25 M_\odot$, $D \approx 40 R_\odot$, $v_w \approx 2\,(GM_2/R_2)^{1/2} \approx 1000$ км$\cdot$с$^{-1}$, $v_{\text{орб}} \approx 300$ км$\cdot$с$^{-1}$, $r_A \approx 1 R_\odot$) величина j будет порядка $j_{\text{крит}}$, так что центробежные силы становятся сильными только вблизи горизонта дыры и аккреционный диск, если он имеется, образуется недалеко от горизонта. Однако этот вывод очень чувствителен к скорости ветра вблизи дыры ($j \sim R_A^2 \sim v_w^{-4}$),

а эта скорость является неопределенной, по крайней мере с точностью до множителя, равного 4. Если скорость ветра меньше, чем ожидается, будет образовываться большой диск; если скорость больше, никакого диска не образуется; если она флуктуирует во времени, диск мог бы образоваться и быть разрушен. Детали см. в работе [181].

В случае быстрого ветра газ, который проходит через ударный фронт, падает квазирадиально на дыру. Он, по-видимому, будет нести с собой какое-то магнитное поле, которое начинается на второй звезде и которое «запутывается» в падающем газе. Эта ситуация напоминает аккрецию массы из межзвездной среды, за исключением того, что скорость аккреции $\dot{M}$ гораздо больше, чем в межзвездном случае. Межарос [129, 130] построил модели для падения, которые похожи на магнитотурбулентные модели [см. уравнение (12) и далее]. Основная разница состоит в том, что поскольку $\dot{M}$ больше, плотность выше, и циклотронное излучение как доминирующий процесс охлаждения заменяется тормозным излучением и/или неупругим комптоновским рассеянием низкоэнергетических циклотронных фотонов высокоэнергетическими электронами. Охлаждение является достаточным, чтобы конкурировать с магнитотурбулентным нагреванием вблизи горизонта, и поскольку большая часть высвобождаемой гравитационной энергии уходит в тепло [см. 13], эффективность генерации излучаемой светимости высока:

$$\varepsilon = L/\dot{M}c^2 \gtrsim 0{,}2. \tag{28}$$

Температура излучающих электронов, регулируемая балансом нагревания и охлаждения, оказывается равной $T \sim 10^9$ К, и излучение в значительной степени идет в диапазон очень жесткого рентгена.

Подходящим выбором свободных параметров в этой модели Межарос достиг разумного согласия с наблюдениями для Cygnus X-1. Однако 1) согласие оказалось не лучше, чем в случае дисковых моделей, 2) в этой модели имеется достаточно *ad hoc* (хотя и разумных) предположений, таких, например, как обсуждавшиеся в связи с (12) и (13), чтобы заставить привередничать (даже) астрофизиков.

Когда вторичная звезда теряет массу за счет истечения из полости Роша или «медленного» ветра, значение j велико по сравнению с $j_{\text{крит}}$ и центробежные силы будут создавать аккреционный диск на достаточно больших радиусах.

По аккреционным дискам было выполнено громадное количество теоретических исследований с момента открытия двойных рентгеновских источников. (Наша стопка препринтов и оттисков составляет семь дюймов публикаций по этому предмету.) Недавний обзор литературы дан Принглом [161], педагогическое введение — в раннем обзоре Новикова и Торна [135].

Теория структуры диска критическим образом зависит от толщины диска: если на радиусе r толщина диска $h \sim r$, диск считается толстым, если $h \ll r$ — тонким. Для тонких дисков теория является вполне приемлемой, для толстых — нет. Лучшие модели для Cygnus X-1 включают диски, которые могут быть толстыми в их внешних областях, но становятся тонкими через несколько e-радиусов по мере движения внутрь, а затем вновь становятся толстыми вблизи дыры.

Толщина диска регулируется балансом между его внутренним давлением p и вертикальным гравитационным сжатием, обусловленным приливным притяжением (тензором Римана) дыры:

$$p/h \approx \rho\,(GM/r^3)\,h. \tag{29}$$

Поскольку плотность внутренней энергии газа (представляющей собой сумму тепловой, магнитной и турбулентной энергий) примерно равна его давлению, то выполняется соотношение

$$\frac{h}{r} \approx \left(\frac{p}{\rho GM/r}\right)^{1/2} = \left(\frac{\text{плотность внутренней энергии}}{\text{плотность гравитационной энергии}}\right)^{1/2}, \tag{30}$$

которое означает, что диск является тонким в том и только в том случае, если он очень эффективно излучает свою внутреннюю энергию.

В тонком диске газ вращается по почти кеплеровским орбитам (неравенство $p \ll GM\rho/r$ подразумевает, что силы радиального давления не могут сильно видоизменить кеплеровское движение). Рассмотрим два смежных кольца газа в таком диске. Внешнее кольцо имеет более низкую кеплеровскую угловую скорость, чем внутреннее. Следовательно, вязкое трение между этими двумя кольцами ускоряет наружное кольцо и замедляет внутреннее кольцо, заставляя наружное кольцо раскручиваться по спирали наружу, в то время как внутреннее закручивается по спирали внутрь. В результате вязкость приводит к отталкиванию между этими двумя кольцами. При рассматривании взаимодействия между всеми кольцами в диске получают результирующее закручивание внутрь, за исключением самых наружных областей. Наружные области, предоставленные самим себе, двигались бы наружу; но они будут подвергаться бомбардировке падающим газом от звезды-компаньона, давая в результате поток, направленный внутрь, детали которого очень трудно анализировать. К счастью, ужасные неопределенности наружных областей не оказывают существенного влияния на двойные рентгеновские модели, поскольку почти весь выход гравитационной энергии и генерация рентгеновского излучения происходят вблизи дыры.

На языке сохранения углового момента вязкие силы уносят угловой момент от газа во внутренних областях диска, тем самым позволяя газу закручиваться внутрь. Угловой момент передается механически (вязкими силами) к внешним областям, где он уносит-

ся плохо понимаемым образом в результате взаимодействия с компаньоном и с втекающей материей.

Предположим, что в кольце $r_1<r<r_2$ вязкость слишком низка, чтобы уносить угловой момент с требуемой скоростью. Тогда газ будет удерживаться в этом кольце при r_2 более сильным напряжением снаружи него, но кольцо не позволит газу проходить внутрь к дыре при r_1. Газ будет накапливаться в кольце. В большинстве случаев этот добавочный газ будет увеличивать вязкие напряжения в кольце, пока они не станут достаточными, чтобы управлять переносом массы. Таким способом будет достигаться стационарное состояние. (Математические детали см. в работах [109, 110] и [118].) Однако в некоторых необычных ситуациях увеличение плотности газа будет приводить к уменьшению вязкости, тем самым заставляя плотность в кольце быстро нарастать, тогда как плотность в соседних областях падает до нуля. Кажется вероятным, что эта «фрагментированно кольцевая» неустойчивость встречается в тонких аккреционных дисках, как только внутреннее радиационное давление $p_{рад}={}^1/_3aT^4$ превышает давление газа $p_{газ}=\mathcal{R}\rho T$ (см. [111]).

Другая неустойчивость, которая развивается быстрее, чем описанная выше, и, возможно, всегда сопровождает ее,— «тепловое нарастание» [160, 163]. Эта неустойчивость включается вязким нагреванием газа, которое с необходимостью сопровождает вязкий перенос углового момента. Когда радиационное охлаждение локально перестает компенсировать вязкое нагревание, температура газа повышается. Если более высокая температура приводит к уменьшению нагревания или к увеличению излучательной способности, температура будет падать до установления равновесного состояния и диск окажется термически стабильным. Но если повышение температуры увеличивает несоответствие между нагреванием и охлаждением, то развивается тепловая неустойчивость.

В действительности нет вполне четкого различия между этими двумя неустойчивостями. Они проанализированы единым образом наряду с другими в окончательной работе Шакуры и Сюняева [178].

В настоящее время наиболее популярные модели для Cygnus X-1 (например, [182]) включают тонкий аккреционный диск, в котором вязкость создается турбулентностью и/или магнитными натяжениями. В сущности они являются дисковыми аналогами магнитотурбулентных моделей радиального втекания Межароса; в обоих случаях нагревание происходит через перезамыкание силовых линий и турбулентную вязкость. В этих моделях диск является оптически толстым, физически тонким и весьма холодным ($T_{поверхн}<<10^7$ К) вплоть до ~30 шварцшильдовских радиусов. В этой точке растущее давление излучения включает обе неустойчивости, приводя самые внутренние части диска в горячее ($T\gtrsim10^9$К), толстое ($h\sim r$), предельно оптически тонкое состояние. Бо́льшая часть наблюдаемого рентгеновского излучения производится в этой внут-

ренней области в результате неупругого комптоновского рассеяния низкоэнергетических фотонов.

Структуры с толстым диском могут возникать из гигантских («сверхэддингтоновских») скоростей аккреции, а также из неустойчивостей, и даже в случае тонкого диска магнитогидродинамические процессы могут генерировать толстые, горячие короны над дисками. Физические процессы в толстом диске могут в свою очередь генерировать сильный ветер, который приводит к потере массы из внутренних областей диска. Предварительные попытки анализировать сверхэддингтоновскую аккрецию, короны и ветер, осуществлены Бисноватым-Коганом и Блинниковым [21], Лиангом и Прайсом [108], Пираном [157], Икке [87, 88] и Бардином [11]. Мы вернемся к этим вопросам в разд. 6 в связи с черными дырами в галактических ядрах и квазарах.

Исследование аккреционных дисков продолжается в настоящее время в ускоренном темпе, и не потому, что имеется огромный прогресс (его нет), а потому, что астрономы недавно поняли, что аккреционные диски представляют собой весьма общее явление во Вселенной: они определенно встречаются вокруг нейтронных звезд и, возможно, вокруг черных дыр в рентгеновских двойных; они могут встречаться вокруг белых карликов в карликовых новых и в старых новых [139, 187, 188]; они могут располагаться вокруг протозвезд в системах T Tauri [118], и они, вероятно, с таким же успехом могут существовать где-нибудь еще.

4.5. НАБЛЮДАТЕЛЬНЫЙ ПОИСК ЧЕРНЫХ ДЫР В ДВОЙНЫХ СИСТЕМАХ

В 1965 г., задолго до открытия рентгеновских двойных, Зельдович и Гусейнов [229] предприняли в каталогах двойных поиски систем, которые могли бы содержать черные дыры. Вероятными кандидатами были спектроскопические двойные с большими массами вторичных звезд, т. е. системы, в которых виден только свет от одной звезды, причем спектральные линии этой звезды испытывают доплеровское смещение, которое периодически осциллирует во времени, и из этих осцилляций можно сделать заключение о большой массе объекта, вокруг которого звезда обращается. К сожалению, все из возможных кандидатов Зельдовича — Гусейнова и все из кандидатов, найденных позже Тримблом и Торном [207], могли бы очень легко быть объяснены без обращения к черным дырам. Пытаясь исключить другие объяснения, Абт и Леви [3] наблюдали половину (пять) из лучших кандидатов Тримбла — Торна и открыли, что четыре из них неправильно классифицированы: они не были двойными вообще!

Это иллюстрирует значительные трудности, с которыми приходится сталкиваться в любом поиске чернодырных двойных в отсутствие какого-либо эффектного признака, такого, как рентгенов-

ское излучение. Из многих кандидатов, которые были предложены, например ε Aurigae [35], β Lyrae [74] и HD 108 [18], ни один не является убедительным.

В противоположность этому весьма убедительные выводы могут быть сделаны для Cygnus X-1 и обнадеживающие выводы — для Circinus X-1. Cygnus X-1 виден в оптический телескоп как голубой сверхгигант спектрального типа O9. 7 Iab. Длина волны линий поглощения осциллирует с достаточно большой амплитудой, указывая, что звезда находится на орбите вокруг массивного компаньона с орбитальным периодом 5—6 дней. Видны также эмиссионные линии с заметно различными осцилляциями. Они убедительно интерпретируются как свет от газа, который течет от супергиганта к его компаньону. Нет никаких данных о свете от компаньона. Результаты этих наблюдений суммированы Болтоном [29].

В рентгеновский телескоп наблюдается рентгеновская светимость с приблизительно равным выходом энергии в диапазонах 1 кэВ$<E<$10 кэВ и 10 кэВ$<E<$100 кэВ и с заметным, но меньшим выходом энергии выше 100 кэВ. Рентгеновское излучение самой низкой энергии (наименьшей проникающей способности) иногда частично затмевается, и эти затмения, как правило, происходят, когда сверхгигант оказывается между нами и своим компаньоном. Можно заключить с убедительностью, что рентгеновское излучение приходит из окрестности компаньона, что линия зрения на Землю проходит выше компаньона (неполное затмение) и что рентгеновский источник время от времени частично затмевается массой, текущей от супергиганта к компаньону. Рентгеновское излучение флуктуирует по интенсивности в масштабах $\geqslant$0,1 с. (Данных еще не вполне достаточно, чтобы убедительно продемонстрировать или исключить более быстрые флуктуации.) Это означает, что в любой момент времени основное количество энергии рентгеновского излучения приходит из областей с размером $\leqslant$(0,1 свет. с)$\approx 3\cdot 10^4$ км. Болтон [29] суммирует эти наблюдения и дает соответствующие ссылки.

Единственное разумное объяснение всего этого — рентгеновское излучение газа, падающего на компаньон размером $\leqslant 3\cdot 10^4$ км, который, согласно существующей теории, может быть белым карликом, нейтронной звездой или черной дырой. (О маловероятной, но мыслимой возможности альтернативного объяснения см. § V работы [29] и имеющиеся там ссылки.) Очевидный путь выбрать тип компаньона— взвесить его.

Существует много различных способов взвешивания. Один из них включает в себя комбинирование законов Кеплера и элементарных геометрических рассмотрений со следующими данными:

А. Орбитальная скорость сверхгиганта, определяемая из осцилляций спектральных линий.

Б. Оценка радиуса сверхгиганта, которая следует из

1) его видимой яркости в оптическом диапазоне (эрг·см$^{-2}$·с$^{-1}$ на Земле);

2) его расстояния, выводимого из

а) степени покраснения его света при прохождении через межзвездную пыль (покраснения, определяемого сравнением наблюдаемых цветов его непрерывного спектра с собственными цветами, создаваемыми атмосферами, которые имеют температуру, плотность и поверхностную гравитацию, выведенные из наблюдаемых спектральных линий);

б) кривой покраснения в зависимости от расстояния для других звезд вблизи луча зрения на Cygnus X-1;

в) абсолютного потока (эрг·см$^{-2}$·с$^{-1}$) видимого излучения с поверхности супергиганта, выводимого из атмосферной температуры, плотности и поверхностной гравитации.

В. Отсутствие рентгеновских затмений, которое устанавливает предельное значение угла наклонения между нашим лучом зрения и орбитальной плоскостью.

Другой метод взвешивания состоит в определении степени деформации супергиганта притяжением компаньона, которую можно вывести из малого изменения видимой яркости с периодом 2,8 дня. Третий (несколько рискованный) способ использует массу сверхгиганта, оцениваемую по его спектральному типу. К счастью, все методы дают согласующиеся результаты (см. [29]): масса темного компаньона лежит между $8M_{\odot}$ и $18M_{\odot}$; она выше $M^{n*}_{\text{макс}}$ и $M^{wd}_{\text{макс}}$; компаньон может быть только черной дырой.

В обычной ситуации астрономы уверенно приняли бы этот результат. Но поскольку в данном случае решается судьба первого открытия человеком черной дыры и поскольку твердые заключения иногда разрушаются своевременно не замеченными систематическими ошибками, астрономы проявляют осмотрительность. Пока не будет найдено дополнительное, независимое подтверждающее доказательство — доказательство скорее положительное, чем отрицательное типа «чем же еще это может быть?» — они не хотят делать вывод, что Cygnus X-1 — действительно черная дыра.

Текущие надежды на подтверждающее доказательство фокусируются на быстрой переменности рентгена. Сюняев [195] указал на то, что если в аккрецирующем газе образуются горячие пятна, то можно ожидать квазипериодических флуктуаций рентгеновского излучения в масштабах, равных орбитальным периодам этих горячих пятен (несколько миллисекунд вблизи дыры). Намеки на такие флуктуации наблюдались с помощью больших рентгеновских телескопов, установленных на ракетах [171, 172], но статистика пока недостаточна, и флуктуаций в действительности могло и не быть [218]. Если будущие наблюдения с большими телескопами на орбите обнаружат миллисекундные флуктуации и если они покажут резкое низкопериодное обрезание, тогда доводы в пользу черной дыры могут стать решающими; и из этого обрезания можно будет

извлечь значение периода последней стабильной круговой орбиты вокруг дыры. Но пока это только надежды.

Из дюжины других более или менее хорошо изученных рентгеновских двойных только одна рассматривается сегодня как вероятный кандидат в черные дыры: Circinus X-1. К сожалению, в этом случае ожидания целиком основаны на поразительном сходстве между ее рентгеновским спектром и коротковременной переменностью и этими же характеристиками для Cygnus X-1 [31, 63, 140]. Только недавно были отождествлены оптическая и радиозвезды, связанные с Circinus X-1 [222]. Они еще недостаточно хорошо изучены, чтобы дать полную и полезную информацию о ее шансах оказаться черной дырой.

5. ЧЕРНЫЕ ДЫРЫ В ШАРОВЫХ СКОПЛЕНИЯХ

5.1. ТЕОРИЯ ЭВОЛЮЦИИ ЗВЕЗДНЫХ СКОПЛЕНИЙ

В разд. 3 мы описывали пути старения и смерти массивных сферических звезд.

Старение звезды характеризуется постепенным сжатием ядра и расширением оболочки. Смерть начинается, по-видимому, с коллапса ядра, ведущего к образованию черной дыры.

Теория предсказывает похожую эволюцию для сферического звездного скопления. В его ядре звезды стремятся эволюционировать к максвелловскому распределению скоростей в результате двухчастичных столкновений звезд, движущихся по гиперболическим орбитам. Однако звезды, лежащие в высокоэнергетическом хвосте этого распределения, ускользают из ядра на слабосвязанные орбиты с большими r или вообще ускользают из скопления («испарение»). В результате ядро постепенно теряет энергию и сжимается, и развивается гало из слабо связанных, радиально выходящих на орбиту звезд. В конечном итоге — если игнорировать трехчастичные столкновения и физические лобовые столкновения звезд — ядро становится неустойчивым вследствие общерелятивистских эффектов и коллапсирует, образуя сверхмассивную черную дыру [90, 231]. Учет трехчастичных столкновений и физических столкновений делает картину более сложной и трудной для анализа, но образование сверхмассивной черной дыры является разумной возможностью; см. § VI обзора Лайтмана и Шапиро [114] и имеющиеся там ссылки.

Характерное время заметного сжатия ядра называется «центральным временем релаксации» скопления. Оно численно равно времени, необходимому для того, чтобы за счет двухчастичных столкновений вблизи центра скопления существенно изменялась орбита звезды. Прицельный параметр b для одиночного столкновения, меняющего орбиту, определяется соотношением

$$G\overline{m}/b = \frac{1}{2}\overline{v^2}, \tag{31}$$

где $\bar{v}$ — среднеквадратичная скорость в ядре, $\bar{m}$ — средняя масса звезды. Здесь, как и в физике плазмы, малоугловые отклонения дают эффект случайных блужданий, который логарифмически больше, чем одиночные столкновения. Следовательно, центральное время релаксации есть

$$t_{rc} \approx \frac{2\,(3/2)^{3/2}}{n_c\,(\pi b^2)\,(\bar{v})\,\ln(0{,}5\,N)} = \frac{(3/2)^{3/2}\,\bar{v}^3}{2\pi G^2 m^2 n_c \ln(0{,}5\,N)} =$$

$$= (1\cdot 10^{13}\ \text{лет})\,\frac{(\bar{v}/10\ \text{км}\cdot\text{с}^{-1})^3}{(n_c/\text{пс}^3)\,(m/M_\odot)^2\,\ln(0{,}5\,N)}\,. \quad (32)$$

Здесь n_c — центральная звездная плотность (число на кубический парсек), а N — полное число звезд в скоплении. Численные и аналитические исследования показывают, что коллапс ядра скопления происходит примерно за 100 t_{rc} [192] и большинство звезд, более массивных, чем $2\,\bar{m}$, оседают в центре скопления через $\sim 30 t_{rc}$ [191].

Характерные времена $100\,t_{rc}$ и $30\,t_{rc}$ больше возраста вселенной для всех типов скоплений, кроме двух: шаровых скоплений и (возможно) центральных ядер некоторых галактик, которые мы рассмотрим соответственно в п. 5.2 и разд. 6.

5.2. НАБЛЮДАЕМЫЕ СВОЙСТВА ШАРОВЫХ СКОПЛЕНИЙ

Шаровые скопления являются сферическими и содержат в типичных случаях от 10^4 до 10^6 звезд в радиусе скопления от 50 до 100 пс; их структура включает центральное ядро радиусом в несколько парсек и плотностью звезд $n_c \sim 10^3$—10^5 пс$^{-3}$, окруженное более разреженным гало. Астрономы могут разрешать индивидуальные звезды в скоплении и из наблюдаемой дисперсии радиальных скоростей, наблюдаемой центральной звездной плотности и оцениваемой средней звездной массы могут вычислять центральное время релаксации скопления. Эти времена заключены в интервале от $\sim 10^7$ до 10^{10} лет (см., например, рис. 1 в работе [10]). Для сравнения фактический возраст всех шаровых скоплений в нашей Галактике (выведенный из возрастов звезд, которые именно сейчас становятся гигантами) равен $1\cdot 10^{10}$ лет, т. е. возрасту самой Галактики. Это с уверенностью наводит на мысль, что значительная часть шаровых скоплений, с которыми была рождена наша Галактика, к настоящему времени умерли в результате коллапса ядра, что примерно 50 или больше из оставшихся ~ 200 умрут в следующие 10^9 лет и что в этих 50 обреченных скоплениях все звезды более тяжелые, чем $2M_\odot$, к настоящему времени осели в центре [115].

Где находятся умершие скопления? Никто не знает. И мы не знаем, на что они должны быть похожи, за исключением того факта, что они могли бы содержать в своих сколлапсировавших ядрах умеренно массивные черные дыры. Сколь массивные? Аргументы, касающиеся влияния трехчастичных столкновений и физических звездных столкновений в умирающем ядре, подсказывают, что

$M_{дыры} \leqslant 1000\ M_{\odot}$ и что бо́льшая часть остальной массы скопления может уже испариться [112].

А какова ситуация с обреченными, но живыми скоплениями, чьи массивные звезды уже должны были осесть в центре? Этот вопрос в последние годы побудил астрономов тщательно исследовать ядра шаровых скоплений. Одним из признаков центральной конденсации, которая могла бы быть теперь массивной черной дырой, должно быть ее гравитационное влияние на звезды в скоплении. Дыра должна преимущественно притягивать звезды к себе, создавая пик в наблюдаемой поверхностной яркости скопления. Хотя несколько наблюдаемых шаровых скоплений имеют неразрешаемые ядра, угловое разрешение все еще недостаточно, чтобы установить или опровергнуть существование в центре черной дыры. Лучшее, что можно сделать,— это установить верхний предел $\sim 10^4\ M_{\odot}$ на массы таких дыр [9, 89].

В 1974 г., когда вопросы умирания скопления находились на ранних стадиях изучения, в шаровых скоплениях были неожиданно открыты рентгеновские источники. К настоящему времени семь рентгеновских источников располагаются в боксе ошибок, который содержит шаровое скопление, и большинство из семи, по-видимому, находится в этих скоплениях. Это означает, что тогда как шаровые скопления содержат только $\sim 10^{-4}$ галактической звездной массы, они содержат несколько процентов ее компактных рентгеновских источников. Более того, все, кроме одного, из семи рентгеновски излучающих шаровых скоплений входят в список обреченных, для которых $100\ t_{rc} < 10^{10}$ лет.

Когда эти факты начали выясняться, казалось очевидным следующее заключение [10, 186]: рентгеновские источники являются сверхмассивными черными дырами ($M \sim 100 \div 10^4 M_{\odot}$), образовавшимися в результате оседания тяжелых ($M \gtrsim 2 M_{\odot}$) звезд и дыр в центре скопления, и рентгеновское излучение генерируется при аккреции межзвездного газа скопления.

Хотя эта теория остается неопровергнутой до наших дней, она перестала быть привилегированной. Главная причина состоит в том, что некоторые из рентгеновских источников в шаровых скоплениях оказались «барстерами» (т. е. излучателями цепочек нерегулярно следующих во времени вспышек), и большинство теоретиков сегодня подозревают, что такая вспышечная активность связана с аккрецией газа на магнитную нейтронную звезду (см., например, [104]). С другой стороны, капризы астрофизической моды могут вновь выдвинуть на передний план модели с черными дырами еще до окончательного установления истины.

5.3. ПРИЛИВНОЕ ВОЗДЕЙСТВИЕ НА ШАРОВЫЕ СКОПЛЕНИЯ

Когда шаровое скопление проходит через диск нашей Галактики, его сильное гравитационное поле производит возмущение в рас-

пределении звезд диска. Это возмущение включает в себя увеличение плотности за скоплением, которое создает гравитационное увлечение («приливное трение») движения скопления. Тремейн [205] показал, что приливное трение достаточно сильно, чтобы увлекать более массивные из шаровых галактических скоплений в центр Галактики за время, меньшее чем 10^{10} лет. Случилось ли это уже с большим числом шаровых скоплений, остается спорным вопросом [203]. Однако если это случилось и если некоторые из этих скоплений содержат сколлапсировавшие ядра, это побуждает к размышлениям о наблюдательных последствиях и о судьбах дыр с массами $(10^2-10^4)M_\odot$, курсирующих в самых внутренних размером в несколько сотен парсек областях Галактики.

6. ЧЕРНЫЕ ДЫРЫ В КВАЗАРАХ И ЯДРАХ АКТИВНЫХ ГАЛАКТИК

6.1. ВВЕДЕНИЕ

Как уже отмечалось в разд. 1, многие скептики смотрят на черную дыру с большим подозрением: как на легкую, бросающуюся в глаза панацею, прописываемую без долгих размышлений против любой астрономической болезни. В случае квазаров (впредь в этой статье термин «квазар» мы будем также употреблять в широком смысле, включающем любое галактическое ядро, обнаруживающее признаки не звездной активности) хорошо бы напомнить, что очень скоро после их открытия [177] они были интерпретированы как мощные примеры гравитационного коллапса (см. в особенности труды Первого Техасского симпозиума 1963 г., а также работу [170]). Вначале черные дыры [117, 174, 228] обсуждались в качестве источников энергии для квазаров наряду со сверхмассивными звездами [83] (называемыми также спинарами [134] или магнетоидами [143]) и плотными звездными скоплениями (см., например, [6, 51]), а также большим количестом короткоживущих спекуляций. Причины, по которым черные дыры только недавно стали самыми популярными кандидатами на роль «основного двигателя» квазаров, заключаются частично в результатах наблюдений, а частично в теоретических рассмотрениях. Исторически теоретическая проблема состояла просто в том, что, несмотря на достаточную глубину потенциала, фактическая эффективность извлечения *полезной* энергии из газа, аккрецируемого массивной ($\sim 10^8 M_\odot$) черной дырой, была первоначально найдена чрезвычайно низкой. Черные дыры тогда более охотно связывали с мертвыми квазарами [117], чем с квазарами в расцвете сил, особенно потому, что полный гравитационный коллапс казался неизбежным конечным состоянием для большинства моделей квазара. Мы докажем ниже, что на самом деле высокая эффективность может быть достигнута вполне правдоподобно.

Детальные обсуждения наблюдений и теорий квазаров содержатся в докладах Восьмого Техасского симпозиума по релятивистской астрофизике[1]) и в Трудах конференций, состоявшихся в Копенгагене и в Кембридже летом 1977 г.[2]) (см. также [80]). В наши планы не входит повторение этих дискуссий; вместо этого мы сконцентрируем наше внимание на кратком обзоре наблюдательных свойств квазаров, делая акцент на те из них, которые прямо указывают на природу главного источника энергии.

6.2. РАДИОСВОЙСТВА

Прошло уже около 30 лет с тех пор, как Дженнисону и Дас Гупта [95] удалось разрешить радиоисточник Cygnus A на две компоненты сравнимой яркости, разнесенные примерно на 100 кпс и находящиеся на противоположных сторонах объекта, оказавшегося гигантской эллиптической галактикой с активным ядром. Это наблюдение позволяло освободиться от использования выдвинутой ранее идеи [8], что экстрагалактические радиоисточники представляют собой сталкивающиеся галактики. Это было также первым примером геометрии источника (двойная структура или в более общем случае линейная структура), которая с тех пор повторялась в радионаблюдениях сильных источников. В самом большом из известных источников, 3C 236 [223], который имеет 6 Мпс в длину, выстроенная двойная структура наблюдается в масштабах от 6 Мпс вплоть до $\leqslant 200$ пс[3]), причем самые малые структуры расположены в ядре ассоциированной галактики. Сейчас имеется несколько примеров очень малых ($\leqslant 1$ пс в поперечнике) компактных источников с тем же позиционным углом в небе, что и бóльшие ($\geqslant 100$ кпс) ассоциированные протяженные источники (см., например, [165]), хотя это ни в коей мере не является общим случаем. Многие компактные источники разрешаются в две компоненты, которые кажутся удаляющимися друг от друга со скоростями, превышающими $2c$ (что объясняется, вероятно, кинематическими эффектами; см., например, [27]). Здесь снова последовательные вспышки происходят обычно в одном и том же направлении (см., например [50]).

Аргументы, независимые от этих наблюдений (см. обзор Де Янга [55]), наводят на мысль, что мощность, питающая большие источники, должна непрерывно генерироваться ассоциированными квазарами и, вероятно, поставляется сквозь окружающую среду по двум каналам. Очевидно, что для центральных областей квазара должен существовать какой-то способ «помнить» определенное направление в течение времени жизни радиоисточника (типичное время 10^6—10^8 лет). Лучший способ запоминания этого — существование спиновой оси, возможно, связанной с центральным плотным

[1]) Annals of the New York Academy of Sciences, 1977.
[2]) Physica Scripta, 17, 1977.
[3]) *R. T. Schillizi*, частное сообщение.

звездным скоплением или газовым облаком, но более вероятно принадлежащей единственному когерентному объекту (черной дыре?), размеры которого гораздо меньше парсека.

Второй аргумент в пользу существования очень плотной структуры в квазарах основывается на наблюдаемом временнóм масштабе вариаций *оптической* светимости компактных «центральных» источников, который в некоторых квазарах достигает ~ 1 дня (см., например, [214]). В отсутствие кинематических эффектов СТО причинность ограничивает размер излучающей области величиной $\lesssim 3 \cdot 10^{15}$ см $\sim 100 (M/10^8 M_\odot)$ шварцшильдовских радиусов. Как следует из энергетических соображений, квазары должны поставлять характерную энергию $> 10^7 M_\odot c^2$ за их времена жизни, а это, вероятно, означает, что массы их источников энергии превосходят $10^8 M_\odot$. Если быстро флуктуирующие квазары типичны для класса как целого, то энергия должна производиться в области, очень близкой к полному коллапсу. (Можно было бы обойти этот аргумент. Если, например, квазар заключает в себе плотное звездное скопление и каждая вспышка соответствует образованию новой нейтронной звезды или черной дыры, тогда скопление может простираться дальше $\gtrsim 1$ пс. Постоянная ось вращения могла бы обеспечиваться либо образованием всех звезд из дифференциально вращающегося диска с параллельными спинами, либо (что более правдоподобно) использованием асимметрии в облаке окружающего газа для формирования выхода энергии. Одна из проблем, возникающих при развитии идеи такого типа, состоит в том, что энергия, связанная со вспышкой индивидуального квазара, по-видимому, не принимает специфических значений, которые могут быть связаны с объектами в $1 M_\odot$; фактически она изменяется в пределах от $\lesssim 10^{-4} M_\odot c^2$ до $\gtrsim 10^3 M_\odot c^2$).

Если квазары обеспечиваются горючим за счет аккреции газа на черные дыры, имеется характерное время [174], которое не зависит от массы дыры: когда эддингтоновская светимость $L_{\text{эдд}} \sim$ $\sim 10^{46} (M/10^8 M_\odot)$ эрг$\cdot$с$^{-1}$ высвобождается при падении газа с (типичной) эффективностью $\sim 0,1$, т. е. когда $L_{\text{эдд}} \sim 0,1 \dot{M} c^2$, то $M/\dot{M} \sim$ $\sim 5 \cdot 10^7$ лет. Это время отличается меньше чем на порядок величины от других независимых оценок времен жизни квазаров.

Переменное *радио*излучение от большинства квазаров, вероятно, является синхротронным излучением, испускаемым релятивистскими электронами, вращающимися в магнитостатических полях. Теория синхротронного излучения предсказывает для радиокомпонент минимальный линейный размер $l_{R,\text{мин}}$ (см., например, [33] и ссылки там). Это предсказание основывается на следующем. Энергия релятивистского электрона, излучающего на частоте ν в магнитном поле с напряженностью B, равна $\gamma mc^2 \sim [2\pi\nu/(eB/mc)]^{1/2} mc^2$. Исходя из термодинамики, мы ожидаем, что энергия электрона превышает произведение постоянной Больцмана и «яркостной температуры» T_B наблюдаемого излучения, а (по определению, пригодному для рэлей-джинсовского участка спектра чернотельного

излучения) $T_B \sim l_R^{-2} L_R \nu^{-3}$, где L_R — полная радиосветимость и l_R — линейный размер источника. Поэтому для данного поля излучения мы имеем нижнюю границу линейного размера источника $l_R \gtrsim l_{R,\text{мин}} \sim B^{1/4}$. Казалось бы, величину l_R можно сделать сколь угодно малой просто путем уменьшения B. Однако это не так, поскольку, если источник становится слишком компактным, плотность энергии радиоизлучения $L_R/4\pi l_R^2 c$ превосходит плотность магнитной энергии $B^2/8\pi$; релятивистские электроны теряют гораздо больше энергии за счет «обратного комптоновского» рассеяния радиофотонов, чем за счет синхротронного «рассеяния» виртуальных квантов в магнитном поле, и требования к полной мощности источника становятся запрещающими. (Каждое обратное комптоновское рассеяние фотона увеличивает его энергию в $\sim\gamma^2$ раз.) Грубая прикидка, которая в большинстве случаев достаточно хороша, показывает, что радиояркостная температура едва ли превышает 10^{12} К и что соответственно $l_{R,\text{мин}} \sim 1$ пс [98].

Тогда размер радиоисточника в типичном случае в $\gtrsim 10^4$ раз больше, чем размер предполагаемой центральной черной дыры. Энергия, необходимая для ускорения излучающих электронов, должна быть перенесена от дыры к компонентам в сравнительно полезной «низкоэнтропийной» форме, такой, как большое количество кинетической энергии. Диссипация всей энергии вблизи горизонта в виде свободно уходящих фотонов или релятивистских частиц невыгодна, если вы хотите объяснить радионаблюдения. Это важное ограничение на модели квазаров с черными дырами.

6.3. ОПТИЧЕСКОЕ И РЕНТГЕНОВСКОЕ ИЗЛУЧЕНИЕ

Типичный квазар излучает в непрерывном спектре полную мощность $\sim 10^{45}$—10^{48} эрг·с$^{-1}$. Оптический спектральный индекс, определяемый как $\alpha = -d\log S_\nu / d\log\nu$, где S_ν — поток энергии на единицу частоты, лежит в интервале $0 \leqslant \alpha \leqslant 4$, и бо́льшая часть мощности может приходиться либо на инфракрасную ($\alpha < 1$), либо на ультрафиолетовую ($\alpha > 1$) область спектра. (Ультрафиолетовый поток, вероятно, ответствен за фотоионизацию областей эмиссионных линий, из которых мы получаем красные смещения и, следовательно, расстояния до квазаров.) Если сейфертовские галактики являются надежным ориентиром, во многих квазарах большое количество возникающей мощности может попадать даже в жесткую рентгеновскую область. (Рентгеновские телескопы пока недостаточно чувствительны, чтобы определить, так ли это в общем случае.)

Для частот, лежащих выше радиодиапазона, ни в коей мере не ясно, что́ является основным процессом излучения, и ввиду крайнего разнообразия наблюдаемых спектров не столь уж неразумно ожидать, что может существовать одновременно несколько конкурирующих механизмов. Наиболее вероятными процессами явля-

ются синхротронное излучение, обратное комптоновское рассеяние (упомянутое выше), нерелятивистское комптоновское рассеяние и тормозное излучение. В нерелятивистском комптоновском рассеянии (см., например, [97]) частота фотона может быть сдвинута вверх на относительную среднюю величину $\Delta\nu/\nu \sim (kT_e/mc^2)$ в каждом акте рассеяния электрона с температурой T_e (в предположении, что комптоновским понижением частоты можно пренебречь, т. е. $kT_e \gg h\nu$). Если источник является оптически толстым по отношению к комптоновскому рассеянию, т. е. если плотность электронов n_e превышает произведение (томпсоновское сечение)$^{-1}$ · (размер источника)$^{-1}$, то отдельный фотон может быть рассеян несколько раз на пути из области источника и может при некоторых обстоятельствах увеличить свою энергию более чем вдвое. Тормозное излучение горячих, свободных электронов, сталкивающихся с ионами (подробную дискуссию см. в работе [135]), будет преобладающим, когда плотность газа вблизи дыры высока. Рентгеновские линии, излучаемые высокоионизованными ионами железа, будут, вероятно, связаны с тепловым тормозным излучением. Если такие линии существуют, они могут быть обнаружены в ближайшем будущем рентгеновскими спутниками типа HEAOB.

Оптическое непрерывное излучение иногда поляризовано, в частности, в источниках, проявляющих быструю переменность. Степень поляризации может в некоторых случаях достигать 30%. Направление вектора поляризации может меняться со временем даже быстрее, чем полная интенсивность. Наибольшая линейная поляризация характерна для синхротронного излучения, хотя величины $\leqslant 10\%$ могут быть достигнуты при комптоновском рассеянии в асферической геометрии.

В большинстве источников излучение также могло бы генерироваться либо нерелятивистскими, либо релятивистскими плазменными механизмами. Даже тот факт, что спектр может простираться на несколько порядков по частоте, не исключает по существу тепловых процессов, ибо если в источнике имеется некоторый интервал температур, то любой спектр с $\alpha > 0$ может генерироваться тепловым образом.

Было сделано много расчетов теоретических спектров, основанных на специфических геометрических или газодинамических допущениях. Характерная черта синхротронных моделей заключается в том, что время жизни оптически излучающих электронов относительно реакции излучения гораздо меньше, чем время, которое требуется свету, чтобы пересечь источник. Это означает, что эти электроны должны быть ускорены на протяжении всего объема излучающей области, и тем самым вновь требуются какие-то средства для переноса мощности от гипотетической центральной черной дыры к области ускорения. Подобные замечания применимы, хотя и менее строго, и к комптоновским моделям.

6.4. АККРЕЦИЯ НА ЧЕРНЫЕ ДЫРЫ

Многие модели оптического и рентгеновского излучения квазаров основаны на квазисферической, или дисковой, аккреции на черные дыры (разд. 4). Важным параметром в этих моделях является отношение времени падения ко времени охлаждения τ_{ff}/τ_c. Если это отношение значительно больше или значительно меньше единицы, то эффективность высвобождения энергии $\varepsilon = L/\dot{M}c^2$ будет низкой и гравитационная энергия аккрецируемого вещества будет заглатываться дырой в виде кинетической или тепловой энергии. Если $\tau_{ff}/\tau_c \sim 1$, то величина ε может быть большой. Для квазисферической аккреции бо́льшая часть падающего газа могла бы иметь форму холодных облаков с низким угловым моментом. Если (в идеальном случае) эти облака сталкиваются очень близко к дыре, где их относительные скорости достигают c, то в облаках будут возникать ударные волны, производящие эффективную диссипацию [61, 167, 168]. (Из наблюдений галактических остатков Сверхновых нам известно, что ударные волны со скоростями 10^{-3}—10^{-2} c достаточно эффективны для ускорения релятивистских электронов и что результирующие эффективности излучения $\varepsilon \gtrsim 0{,}1$ являются вполне правдоподобными, если этот тип столкновения может произойти в действительности.) Как говорилось выше, при дисковой аккреции ε также может быть величиной $\gtrsim 0{,}1$.

Неустойчивости, которые являются бедствием для моделей рентгеновских двойных, в полной мере присутствуют и в дисковых моделях квазаров. Самые внутренние области диска, окружающие черную дыру с массой $\sim 10^8 M_\odot$, аккрецирующую на эддингтоновском пределе, должны иметь температуры $\gtrsim 10^6$ К. Это значит, что отношение давления излучения к газовому давлению (см. разд. 4) велико и что охлаждение в линиях (см., например, [34]) может быть очень важным. Оба эти фактора сильно способствуют тепловой неустойчивости. (В любом случае простые дисковые модели без неустойчивостей не могут воспроизвести полного диапазона наблюдаемого спектра.) Подобная неустойчивость может встретиться, если аккрецируется магнитное поле и синхротронное излучение релятивистских электронов важно во внутренней части диска [163].

Стандартные, не учитывающие неустойчивостей предположения о процессе дисковой аккреции, которые могут быть в какой-то мере справедливыми для рентгеновских двойных, вероятно, не пригодны для квазаров.

В литературе по квазарам обсуждаются два варианта стандартной модели дисковой аккреции. Во-первых, диск может быть достаточно массивным, чтобы нельзя было игнорировать его самогравитацию. (Условием этого является то, что плотность материи в диске на радиусе r превосходит «предел Роша» $\sim M/r^3$, где M — масса дыры [147].) В этом случае угловой момент может переноситься наружу скорее гравитационными взаимодействиями, сходными

с теми, которые действуют в спиральных галактиках, чем локальными вязкими напряжениями. Во-вторых, Абрамович, Яросинский и Сикору [2], а также Козловский, Яросинский и Абрамович [102] исследовали характер эквипотенциальных поверхностей вблизи горизонта событий керровской дыры. Они нашли, что если вязкие и радиационные напряжения не играют существенной роли, то аккрецирующая материя заполнит пространство, ограниченное нулевой эквипотенциальной поверхностью, и что исключенный объем будет определять начало пары каналов, вдоль которых может фокусироваться энергия, необходимая для образования двойного радиоисточника.

В стандартном аккреционном диске энергия выделяется локально в виде излучения, которое свободно уходит с поверхности диска. В случае квазаров более обещающей является скорее та идея, что бóльшая часть энергии освобождается вблизи дыры в неизлучательной форме и что бóльшая часть излучения непрерывного спектра генерируется на некотором расстоянии от дыры, где выходящий поток энергии взаимодействует с окружающим веществом.

В одном случае реализации этой идеи (см. [11, 87, 88, 108, 157]) освобождаемая гравитационная энергия запасается в «короне» над диском. Энергия может быть унесена в виде излучательным или тепловым механизмом управляемого ветра — сдвинутая по масштабу версия солнечного ветра, который уносит бóльшую часть энергии, накопленной в солнечной короне. Были найдены решения подобия, в которых малая часть вещества, аккрецирующего в диске, «принимается» дырой и может генерировать светимость $\sim L_{\text{эдд}}$. Остальное вещество уносится радиационным давлением. При этом оказывается возможным получать потоки, коллимированные параллельно и антипараллельно спиновой оси.

В альтернативной схеме (см. [26], а также статью Блэндорфа в книге [80] и приведенные там ссылки) энергия и угловой момент аккрецирующего газа извлекаются электромагнитными скручивающими усилиями, действующими вблизи дыры. Это на самом деле может быть сделано с достаточно высокой эффективностью даже в осесимметричной геометрии. Рассмотрим магнитное поле, внедренное в диск. В первом приближении поле будет «вморожено» в материю, вращающуюся в диске (вследствие огромной электрической проводимости, что подразумевает «идеальное МГД-условие» $\mathbf{E}+\mathbf{v}\times\mathbf{B}/c=\mathbf{0}$. Ротор этого уравнения подразумевает $\partial\mathbf{B}/\partial t=\nabla\times\times(\mathbf{v}\times\mathbf{B})$, что прямо интерпретируется как вмораживание магнитного поля в вещество). Магнитные силовые линии, выходящие из диска и «вмороженные» во вращающуюся в диске материю, будут генерировать электрическое поле, каким бы его видели локально невращающиеся (стационарные) наблюдатели. Это электрическое поле создает электрическую разность потенциалов поперек самых внутренних частей диска и фактически поперек дыры, точно так же, как в диске Фарадея. Эта разность потенциалов будет заставлять токи

течь вдоль магнитных силовых линий из диска, устанавливая магнитосферу вокруг дыры. В конце концов эти токи будут генерировать тороидальную компоненту магнитного поля, так что силовые линии будут сноситься назад движением вещества. Поэтому будет существовать сопротивляющийся момент вращения, действующий на любое вещество вблизи дыры, и это может приводить к переносу углового момента (и энергии) не наружу в плоскости диска (как в обычных моделях с вязкостью), а перпендикулярно диску в виде электромагнитного или гидромагнитного потока Пойнтинга.

Тот же механизм может вести к извлечению спиновой энергии из самой дыры. Из керровской черной дыры с удельным угловым моментом a в принципе можно извлечь [46] долю энергии $\{1-[[1+(1-a^2/M^2)^{1/2}]/2]^{1/2}\}$ (изменяющуюся от 0 до 29% при возрастании a от 0 до M). Однако чтобы это осуществилось на практике, требуются токи, свободно текущие поперек горизонта. Поскольку частицы должны двигаться внутрь на горизонте и могут, по-видимому, двигаться наружу на больших расстояниях, должен существовать какой-то источник зарядов, переносящих ток во внутренней магнитосфере. Он может обеспечиваться разрушением вакуума над горизонтом, как при ударе молнии. Это приводит к тому, что при ожидаемых внутри ядра квазара условиях существуют простые механизмы, способные произвести это разрушение. Это дает альтернативный способ высвобождения значительной части энергии покоя аккрецируемого вещества. На самом деле любой аккрецирующий замагниченный газ будет, по-видимому, нестабильным, так что бóльшая часть энергии будет выделяться скорее во взрывных вспышках [183], чем в описанных здесь идеализированных стационарных во времени процессах.

Конечным результатом действия как коронального, так и электромагнитного механизма будет, по-видимому, коллимированный ветер, вполне вероятно, движущийся с релятивистской скоростью. Однако он вряд ли может создавать двойной радиоисточник без дополнительных фокусирующих механизмов вдали от дыры.

Если оказывается, что основная часть наблюдаемой энергии излучения имеет вторичное происхождение и не генерируется очень близко к дыре, то становится гораздо труднее отличить модель с черной дырой от модели, основанной, например, на сверхмассивной звезде. Вероятно, в понимании природы излучающей области наибольшие надежды можно возлагать на рентгеновские наблюдения удаленных квазаров и оптические наблюдения «Лацертид» (класс квазароподобных объектов, характеризуемых отсутствием эмиссионных линий и наличием высокой поляризации, крутого спектра и экстремально быстрой переменности).

Если черная дыра в квазаре не окружена особенно массивным диском, она должна получать газообразное топливо со скоростью $\sim(1-100)M_{\odot}$ год$^{-1}$. Постулируются два характерных механизма. Во-первых, плотное скопление звезд вокруг массивной дыры долж-

но иметь в центре пик плотности, обеспечиваемый действием механизмов, обсужденных в разд. 5 для черных дыр в шаровых скоплениях. Звезды внутри этого пика могут разрушаться приливными силами или столкновениями, и значительная часть газообразных продуктов разрушения должна падать на дыру [64, 113]. Если бы дыра была достаточно массивна ($M \gtrsim 10^6 M_\odot$), она могла бы притягивать достаточно плотные области скопления, так чтобы обеспечить топливом даже наиболее яркие квазары [225]. Можно было бы думать, что поставляемая таким образом плазма будет неспособна определять выделенную ось. Напротив, Бардин и Петтерсон [12] показали, что если дыра быстро вращается, то в пределах ~100 шварцшильдовских радиусов аккрецируемая плазма будет оседать в экваториальной плоскости дыры. Интересно, что дыра с массой $M_h \gtrsim 10^8 M_\odot$ создает приливное ускорение $\sim c^6 R_\odot (GM_h)^{-2}$ в звезде типа Солнца, которое *меньше*, чем ее собственная гравитация $\sim GM_\odot R_\odot^{-2}$. Это означает, что звезды не будут разрушаться приливными силами до пересечения ими горизонта, и это может обеспечить способ выключения квазаров, когда их массы становятся больше чем $\sim 10^8 M_\odot$[82].

Другим постулируемым источником аккрецируемого вещества для черной дыры является окружающий газ в окрестности дыры. Такой газ мог бы генерироваться звездными процессами (Сверхновые, планетарные туманности и т. д.) в гипотетической окружающей галактике (см., например, [67]) или выхватываться из проходящей галактики (см., например, статью Ганна в книге [80]).

Образование черной дыры прежде всего дает богатые возможности для произвольных теоретических выводов. Как обсуждается в п. 7.1, дыра может быть первичной и на самом деле играть роль зародыша конденсации окружающей галактики. Однако она может быть также естественным продуктом эволюции массивного ядра звездного скопления, увеличенным в масштабе вариантом сценария для шарового скопления (см. разд. 5). Как и в шаровых скоплениях, дыра может проглотить заметную часть массы первичного скопления, если только имеется эффективный механизм диссипации (например, звездные столкновения), который может увеличить энергию связи звезд. Этот вопрос подробно обсуждается Рисом в работе [80].

6.5. НАБЛЮДЕНИЯ ЧЕРНЫХ ДЫР В КВАЗАРАХ

В качестве заключения разд. 6 отметим, что модель квазара с черной дырой, по-видимому, не даст легко узнаваемых в наблюдениях признаков, как, впрочем, и многие конкурирующие модели. Например, обнаружение значительного смещения в радиоположении от одного всплеска до другого более всего согласовывалось бы с моделью звездного скопления; демонстрация убедительной периодичности в световой кривой квазара [144] могла бы служить сильным свидетельством в пользу несколлапсированной сверхмассивной

звезды. Как и в случае двойных рентгеновских источников, по-видимому, лучшее, на что можно надеяться в ближайшем будущем,— это цепь аргументов, основанная и на теории, и на наблюдениях, которая приведет черную дыру в положение единственного правдоподобного кандидата на роль центрального объекта.

Примеры таких аргументов привели Линден-Белл и Рис [119], исходя из характера излучения, для компактного радиоисточника, помещенного точно в центре нашей собственной Галактики. Позднее Келлерман и др. [100] нашли, что четвертая часть компактного радиоизлучения из галактического центра приходит из области $\sim 10^{14}$ см в поперечнике, что соответствует лишь ~ 100 шварцшильдовским радиусам для черной дыры с массой $3 \cdot 10^6 M_\odot$. (Из наблюдений узкой инфракрасной линии в области с линейным размером $\leqslant 0{,}1$ пс Вольман и др. [224] смогли показать, что максимальная масса какой-либо центральной дыры не превосходит $4 \cdot 10^6 M_\odot$.) Как показали Фабиан и др. [61], относительно неимпозантные объекты, подобные Centaurus A, также могли бы содержать в центре черную дыру. В этих случаях наблюдаемая мощность ограничивается, возможно, снабжением газом, которое должно быть гораздо меньше, чем эддингтоновский предел.

Наконец, Сарджент и др. [175] и Янг и др. [226] исследовали ядро эллиптической галактики M87. Из фотометрических и спектроскопических измерений они нашли, что в центральной области ~ 100 пс масса $\sim 5 \cdot 10^9 M_\odot$, по-видимому, в незвездной форме. Если она отождествляется с черной дырой, то интересно, что верхний предел на радиус самой малой компоненты центрального радиоисточника составляет примерно лишь 50 шварцшильдовских радиусов [99]. Может быть, исходя из всего этого, мы уже почти разрешили черную дыру в квазаре.

7. ПЕРВИЧНЫЕ ЧЕРНЫЕ ДЫРЫ

7.1. КОСМОЛОГИЧЕСКОЕ ОБРАЗОВАНИЕ ПЕРВИЧНЫХ ДЫР

Наряду с черными дырами, образующимися в процессе естественной эволюции звезд или галактик, могут существовать также черные дыры, возникшие как первичные образования в самые ранние эпохи космологического времени. В частности, если некоторая часть ранней вселенной была достаточно неоднородной, то массы, гораздо меньшие, чем максимальная масса нейтронной звезды $M^{n*}_{\text{макс}}$, могли бы коллапсировать, образуя черные дыры. Экстремальная форма неоднородной космологии, изучаемая несколькими авторами (см., например, [14, 132, 166]), называется «хаотической». Имеется веская причина, почему вселенная могла не быть изотропной и однородной в самые ранние времена. Она заключается в том, что масса внутри горизонта частицы, т. е. в сущности мера количества вещества в причинном контакте, уменьшается до нуля при подходе

к начальной сингулярности. Однако существующая вселенная оказывается и изотропной, и в достаточно больших масштабах однородной, как следует в особенности из наблюдений изотропии микроволнового фона. Проблемой, сложной для хаотической космологии и до сих пор удовлетворительно не разрешенной, является установление эффективного механизма сглаживания всех этих вещей в некоторую промежуточную эпоху.

Чтобы масса M коллапсировала с образованием черной дыры в ранней высокотемпературной вселенной, ее плотность ρ должна быть примерно вдвое больше плотности $\bar{\rho}$ окружающей ее среды, когда вселенная расширилась достаточно для того, чтобы горизонт частицы (длина $\sim ct$, где t — возраст вселенной) равнялся размеру массы [длина $\sim (M/\bar{\rho})^{1/3}$]. Используя космологическую связь между плотностью и возрастом, $\bar{\rho} \sim (Gt^2)^{-1}$, мы получаем соотношение [77] между массой дыры M, плотностью $\bar{\rho}$ и временем t, когда она образуется:

$$M \sim (c^6/\bar{\rho}G^3)^{1/2} \sim 10\,(\bar{\rho}/10^{15}\ \text{г}\cdot\text{см}^{-3})^{-1/2} M_\odot \sim \sim 10\,(t/10^{-4}\ \text{с})\, M_\odot. \quad (33)$$

Во времена $t \leqslant 10^{-4}$ с, когда температура T превосходит $\sim 10^{12}$ К, космологические представления становятся весьма спекулятивными главным образом из-за незнания правильного уравнения состояния, которое может варьировать от предельно мягкого (см., например, [65, 75]) до предельно жесткого [116, 227]. Даже если бы нам было известно уравнение состояния, незнание спектра флуктуаций не позволяет строить *a priori* правдоподобные догадки относительно характера спектра масс первичных черных дыр (ПЧД). Этот вопрос подробно обсуждается Карром [36]. Некоторые авторы [38, 116, 127] полагают, что минимальная масса ПЧД $\sim 1 M_\odot$. Это исключило бы возможность наблюдаемой плотности дыр, излучающих за счет процесса Хокинга (см. ниже).

Вначале высказывались опасения [230], что ПЧД могли бы нарасти в течение эпохи преобладания излучения до дыр с массой $\sim 10^{16} M_\odot$ к моменту рекомендации электронов, и протонов в водород при температуре $T \sim 4\,000$ К. Карр и Хокинг [40] нашли, однако, что дыра, вероятно, только удвоит свою массу после того момента, когда она оказывается внутри своего горизонта, и что широкий спектр дыр с массами $\ll 10^{16} M_\odot$ может пережить расширение вселенной без значительного роста. Любопытно, что при максимально жестком уравнении состояния, которое приравнивает давление плотности массы-энергии [227] и которое может возникать во времена $t \ll 10^{-4}$ с, этот быстрый рост может происходить и размеры дыр будут не отставать от размера горизонта до тех пор, пока жесткое уравнение состояния справедливо [116].

Значительное образование ПЧД в горячей вселенной встречается с одним серьезным препятствием, которое делает эту идею на

первый взгляд до некоторой степени невероятной [135]. Мы можем характеризовать существующую плотность ПЧД параметром $\Omega_{\text{ПЧД}}$, равным отношению их средней плотности массы к критической плотности $\rho_c = 3H_0^2/8\pi G$ (H_0 — постоянная Хаббла), необходимой для замыкания вселенной. Фотоны (космического фона и в спектральных линиях удаленных квазаров) испытывают красное смещение при расширении вселенной, так что их энергии в эпоху, когда температура фонового излучения была T, удовлетворяют соотношению

$$h\nu = h\nu_0 (T/T_0) = h\nu_0 (1+z)^{-1}, \tag{34}$$

где V_0, T_0 — современные частота и температура фона, z — красное смещение. В противоположность этому энергия в черной дыре не уменьшается при расширении. Это означает, что параметр плотности при красном смещении образования z_f должен удовлетворять неравенству

$$\Omega_{\text{ПЧД}}(z_f) \leqslant 10^4\, \Omega_{\text{ПЧД}}(0)(1+z_f)^{-1}, \tag{35}$$

где мы использовали тот факт, что современный микроволновой фон ответствен за 10^{-4} критической плотности. Поскольку типичное красное смещение образования ПЧД есть $z_f \geqslant 10^{10}$ и поскольку из наблюдений $\Omega_{\text{ПЧД}}(0) \leqslant 1$, мы должны заключить, что $\Omega_{\text{ПЧД}}(z_f) \leqslant 10^{-6}$. Кажется, нет особых причин, почему дыры должны образовываться с плотностью $\Omega_{\text{ПЧД}} \sim 10^{-8}$—$10^{-6}$, чтобы быть наблюдаемыми теперь. В самом деле, если первичные дыры вообще возникали, кажется более вероятным, что они должны были бы возникать с $\Omega_{\text{ПЧД}}\{z_f\} \sim 1$ и потому на много порядков нарушать современные наблюдательные запреты.

Имеется один способ избегнуть этой и связанных с ней трудностей — постулировать холодную хаотическую вселенную, в которой фоновое излучение возникло в достаточно недавние времена, т. е. при $z \ll 10^{10}$. Возможные механизмы «недавней» генерации необходимой энтропии (т. е. фотонов) включают ударные волны [166], аккрецию дырами с $M \geqslant 1 M_\odot$ [14, 38] и испарение дыр с $M \leqslant 10^{15}$ г [37, 78]. Космологии этого типа сталкиваются с серьезными проблемами, когда они пытаются объяснить не только существующую энтропию на барион [41], но также и термализацию микроволнового фона [232] и производство гелия и дейтерия [208, 233]; однако такие космологии не могут быть легко исключены.

Статистические флуктуации в плотности первичных черных дыр не сцеплены с окружающим излучением и поэтому могут нарастать за космологические времена и тем самым являться «зародышами» образования галактик и скоплений галактик. Межарос [127, 128] показал, что «$\sqrt{N}$»-флуктуации ПЧД с $M = 1 M_\odot$ являются достаточными (см., однако, [39]); Риан [173] и Барроу [14] рассмотрели эффекты, связанные с еще большими массами. Ввиду известных трудностей, связанных с объяснением образования галактик в стандартной космологии, жизнеспособность «засеивания» ПЧД заслуживает дальнейшего изучения.

7.2. НАБЛЮДАЕМОСТЬ ПЕРВИЧНЫХ ДЫР

Из изложенного выше должно быть ясно, что теоретические аргументы в пользу большого числа первичных черных дыр чрезвычайно спекулятивны и далеко не бесспорны. Каковы перспективы их обнаружения в будущем и какие пределы можно установить из существующих наблюдений?

Ввиду глубокого смысла, который будет иметь их открытие для теоретической физики, не удивительно, что в последние несколько лет внимание сконцентрировано на свойствах дыр с массами $M \sim 10^{15}$ г, испаряющихся, согласно Хокингу [78, 79], во времена, сравнимые с существующим возрастом Вселенной. Пейдж [148—150] вычислил скорости испускания частиц в процессе Хокинга, включая влияние флуктуаций вращения и заряда дыры. Он нашел значение $5\cdot 10^{14}$ г $\leqslant M_H \leqslant 7\cdot 10^{14}$ г для «массы Хокинга» (т. е. массы дыры, которая испаряется именно в настоящем возрасте Вселенной); точное значение M_H зависит от спина дыры. Он также нашел, что значительная часть ($\geqslant 0{,}2$) испаряемой светимости должна выделяться в виде γ-лучей с пиком при энергии ~ 100 МэВ. Это γ-излучение можно было бы обнаружить тремя способами.

1. Интегральный фон от ПЧД до красного смещения ~ 1 может быть детектируемым. Это может дать верхний предел на плотность ПЧД, которую Пейдж и Хокинг [151] (ср. [45]) оценили как $\bar{n}_H \leqslant 10^4$ пс$^{-3}$ для дыр с массой $\sim M_H$. Это соответствует $\Omega_{\text{ПЧД}}(M_H) \leqslant 10^{-8}$. Сравнение между предсказанным и наблюдаемым спектрами в принципе могло бы дать положительный результат.

2. Можно было бы надеяться обнаружить γ-излучение от отдельных, близко расположенных ПЧД. Если дыры скапливаются около галактик, то локальная плотность n_H может превышать $\bar{n}_H$ на множитель группирования $C \leqslant$ (размер Вселенной)/ (размер нашей Галактики) $\sim 10^6$. Из этого последнего предела мы можем заключить, что не следует ожидать увидеть *отдельный источник* ближе, чем на расстоянии

$$10^{15}\,(\bar{n}_H/10^4)^{-1/3}\,(C/10^6)^{-1/3}\ \text{см} \tag{36}$$

(т. е. это вне Солнечной системы). Поток γ-излучения от источника на расстоянии 10^{15} см составлял бы $\leqslant 10^{-9}$ фотон$\cdot$см$^{-2}\cdot$с$^{-1}$, что полностью ненаблюдаемо.

Мы можем установить верхний предел на дозволяемую локальную плотность более тяжелых дыр, потребовав выполнения неравенства $\Omega_{\text{ПЧД}} \leqslant 1$. Этот предел принимает значение

$$n\,(M) \leqslant 10^{11}\,C(M/10^{15}\,\text{г})^{-1}\ \text{пс}^{-3}. \tag{37}$$

Таким образом, максимальная дозволенная пространственная плотность ПЧД есть $n(M) \leqslant 10^9 C$ пс$^{-3}$ для $M \sim 10^{17}$ г. (Так как маловероятно, что случайная скорость дыры превышает 100 км$\cdot$с$^{-1}$,

шансы столкновения дыры с Землей (см. [92]) в этом столетии в высшей степени незначительны!)

3. Что касается третьей наблюдательной возможности, а именно *взрыва черной дыры*, то перспективы γ-детектирования кажутся наиболее обещающими. Когда масса дыры падает ниже M_H, излучаемая мощность возрастает по крайней мере как $(-t)^{-2/3}$, где время t отсчитывается от момента окончательного исчезновения дыры. В результате последние остатки массы покоя дыры излучаются взрывным образом. Данный тип частиц начинает изучаться, когда шварцшильдовский радиус дыры сокращается до соответствующей комптоновской длины волны. Если, как первоначально предсказывал Хагедорн [75], число видов частиц с массами выше массы пиона возрастает с ростом массы экспоненциально, то энергия $\sim 10^{34}$ эрг будет высвобождена за время $t \gtrsim 10^{-7}$ с главным образом в виде γ-излучения с энергией 250 МэВ [42]. С другой стороны, для более жестких уравнений состояния, совместимых с простейшими кварковыми моделями (более точно, для тех моделей высокотемпературного вещества, которые имеют скорость звука $\gtrsim c/\sqrt{5}$, см. [42]), выход энергии будет далеко не постепенным:

$$E(>t) \sim 10^{29}(-t/1\ c)^{1/3}\ \text{эрг}, \qquad (38)$$

где $E(>t)$ — энергия, излучаемая после момента времени t. В этом случае взаимодействие частиц можно игнорировать, и продукты взрыва должны быть наблюдаемы с энергиями вплоть до планковского значения $\sim(\hbar c/G)^{1/2} \sim 10^{28}$ эВ. Независимо от высокотемпературного поведения вещества ближайший взрыв за время наблюдения в один месяц произошел бы на расстоянии

$$\sim 2\,(\bar{n}_H/10^4\ \text{пс}^{-3})^{-1/3}(C/10^6)^{-1/2}\ \text{пс} \qquad (39)$$

от Земли. Даже в наилучшем случае хагедорновского взрыва и максимальной плотности ПЧД можно ожидать поток только в 10^{-1}—10^{-2} γ-фотонов на 1 см², для чего требуется детектор с хорошим временны́м разрешением и площадью $\gtrsim 100$ см² [42, 151]. Поэтому хотя наблюдение продуктов взрыва от высокотемпературной дыры в принципе представляет уникальный эксперимент в физике сверхвысоких энергий, практические трудности в его осуществлении кажутся труднопреодолимыми, если ПЧД с массой $\sim 10^{15}$ г не имеет максимальной дозволенной плотности.

Более обнадеживающая идея [167[основана на генерации взрывающейся дырой значительного количества электронно-позитронных пар. Эти частицы будут быстро остановлены межзвездным магнитным полем, но при этом поле будет сметено и сжато так, что образуется когерентный низкочастотный импульс электромагнитных волн типа того, который первоначально предлагался Колгейтом и Нёрдлингером [52] для Сверхновой.

Если в лучшем случае энергия порядка половины энергии дыры выделяется в виде пар с лоренц-фактором γ, что соответствует тем-

пературе взрыва $\sim\gamma$МэВ, то должно быть произведено $10^{43}\gamma^{-2}$ пар. При условии, что длительность взрыва Δt достаточно мала, пáры должны образовать тонкую, электрически проводящую оболочку, расширяющуюся радиально в окружающем поле. Пока $\gamma\gg1$, наблюдателю, движущемуся с оболочкой, магнитостатическое поле будет казаться подобным электромагнитному излучению, и виртуальные кванты в этом поле будут отражаться от оболочки. Возвращаясь в систему отсчета взрыва, мы найдем, что поток энергии «рассеянного» излучения пропорционален произведению γ^2 на величину падающего потока магнитостатической энергии (один множитель γ для каждого преобразования Лоренца).

Таким образом, электроны и позитроны затормозятся, когда энергия взрыва $E\sim10^{37}/\gamma^{-1}$ сравняется с излучаемой энергией, т. е. когда оболочка расширится до радиуса R, удовлетворяющего соотношению

$$E\sim\gamma^2\frac{B^2}{8\pi}\cdot\frac{4\pi}{3}R^3, \tag{40}$$

где B — окружающее магнитное поле. Теперь падающие виртуальные фотоны будут иметь длину волны $\sim R$, и, таким образом, дважды сдвинутая по Доплеру наблюдаемая длина волны будет равна

$$\lambda\sim\gamma^{-2}R\sim(6E/\gamma^8B^2)^{1/3}. \tag{41}$$

Для того чтобы произошло подобное взаимодействие, должен быть выполнен ряд физических критериев: взрыв должен быть достаточно быстрым (т. е. $\Delta t\ll R/c\gamma^2$); оболочка должна иметь достаточную поверхностную проводимость; пáры должны иметь возможность избежать аннигиляции в расширяющемся огненном шаре и т. д. Оказывается [25, 167], что возможность обнаружения, вероятно, оптимизируется в случае, если дыра взрывается, когда ее масса $\sim10^{11}$ г, выделяя энергию $\sim10^{32}$ эрг в виде пар с энергиями 100 ГэВ. Такая дыра, помещенная в межзвездное магнитное поле (~3 мкГс) на расстоянии ~10 кпс от центра Галактики, должна дать линейно поляризованный радиоимпульс с потоком энергии $\sim10^{-23}$ эрг$\cdot$см$^{-2}\times$ $\times$Гц$^{-1}$ на частоте $\omega\leqslant1$ ГГц. Для детектирования этого импульса достаточно приборов довольно умеренной чувствительности. Существующий верхний предел на частоту следования взрывов в предположении, что бóльшая часть энергии производится в диапазоне частот 100 МГц—1ГГц, составляет $\sim10^{-5}$ пс$^{-3}\cdot$год$^{-1}$ [125]. Ввиду гораздо большей энергетической чувствительности радиотелескопа по сравнению с γ-детектором не удивительно, что этот предел в $\sim10^{8}$ раз лучше, чем существующий предел для γ-детектора. В принципе направленный поиск с многофазовой антенной и использованием методов уменьшения дисперсии мог бы улучшить этот предел до $\sim10^{-12}$ пс$^{-3}\cdot$год$^{-1}$[125], что соответствует $\Omega_{пчд}(M_H)\sim10^{-14}$. Если дыра взрывается при температуре ~10 ГэВ, то может генерироваться когерентное оптическое излучение [94]. В этом случае

получаются несколько меньшие верхние пределы [94]. Однако следует еще раз подчеркнуть, что из-за неопределенностей в физике взрыва и взаимодействия с окружающей средой эти пределы соответствуют скорее детектируемости при наиболее оптимистических предположениях, чем твердым верхним пределам на $\Omega_{\text{пчд}}(M_H)$.

7.3. ОГРАНИЧЕНИЯ НА ПЛОТНОСТЬ ПЕРВИЧНЫХ ДЫР

Точные верхние пределы на $\Omega_{\text{пчд}}$ для более массивных дыр установить еще труднее, чем для дыр, близких по массе к M_H. На самом деле различными авторами высказывалось предположение, что «недостающая масса», требуемая для стабилизации галактик, связывания скоплений галактик и даже для замыкания вселенной для тех, кто считает это желательным, может принимать эту форму. (Наоборот, если принять аргументы, например, Готта и др. [71], в пользу вселенной с низкой плотностью, где дыры скапливаются, подобно галактикам, тогда очевидно $\Omega_{\text{пчд}} < \Omega_{\text{полн}} \sim 0{,}04$.)

Одно ограничение, действительно справедливое для дыр (и фактически для любого достаточно компактного объекта) в диапазоне масс $10^4\ M_\odot \leqslant M \leqslant 10^{15}\ M_\odot$, описано Прессом и Ганном [159]. Оно опирается на тот факт, что удаленный оптический или радиоисточник может быть сфокусирован с помощью эффекта гравитационной линзы, образуя два похожих изображения, если имеется первичная черная дыра, лежащая между источником и наблюдателем (см., например, [169]). Если мы рассматриваем лучи от источника с красным смещением $z \sim 1$, т. е. на расстоянии $\sim c/H_0$, то расстояние ближайшего подхода лучей, проходящих с одной стороны от дыры, равно $d \sim (GM/c^2\delta\theta) \sim (c\sigma\theta/H_0) \sim (GM/cH_0)^{1/2}$, где $\delta\theta$ — угол, на который изгибается луч. Чтобы получить два изображения сравнимой яркости, два луча должны отклониться на одинаковые углы $\delta\theta$, а это значит, что источник должен лежать позади черной дыры в объеме $\sim (c/H_0)(GM/cH_0)$. Следовательно, вероятность наблюдать двойное изображение, например удаленного квазара, равна примерно $\Omega_{\text{пчд}}$. Ожидаемое угловое разделение равно $\delta\theta \sim 10^{-6}(M/1M_\odot)^{1/2}$ угловых секунд и служит мерой массы дыры. Наименьший угловой размер, измеряемый с использованием межконтинентальной радиоинтерферометрии, $\sim 10^{-4}$ угловых секунд, что соответствует $M \sim 10^4 M_\odot$. На самом деле двойная структура в масштабе 10^{-3} угловых секунд наблюдалась в нескольких удаленных радиоисточниках, но это, как полагают, не связано с эффектом гравитационной линзы [27]. Оптические наблюдения в масштабе угловой секунды показывают, что верхний предел $\Omega_{\text{пчд}}(10^{12}–10^{15} M_\odot) \leqslant 0{,}2$.

Дыры с $M \geqslant 10^{15}\ M_\odot$ могут быть ограничены по динамическим соображениям величиной $\Omega_{\text{пчд}} \leqslant (M/10^{16} M_\odot)^{-2}$ из-за отсутствия наблюдаемого гравитационного влияния на нашу Галактику [159]. Такие же пределы могут быть установлены для дыр в галактических гало и в скоплениях галактик [168, 209]. Если мы готовы де-

лать некоторые допущения о характере межгалактической среды и излучательной способности плазмы вблизи дыры, как обсуждалось в разд. 3, то можно получить некоторые довольно интересные ограничения на $\Omega_{\text{ПЧД}}$ для больших масс [53, 168]. Например, используя формулу (14) с $\xi=10^{-11}$ и $\varepsilon=10^{-1}$, получим предсказываемую плотность энергии в излучении от таких дыр:

$$\sim \rho_c \frac{\Omega_{\text{ПЧД}}}{MH_0} L \sim 10^{-25}\, \Omega_{\text{ПЧД}} \frac{M}{M_\odot} \text{ эрг. см}^{-3}. \qquad (42)$$

Для плотности $\Omega_{\text{ПЧД}}\sim 1$ в дырах с массой $\gtrsim 10^{12} M_\odot$ этот фон, вероятно, мог бы быть обнаружен, в каком бы волновом диапазоне он ни оказался. Подобное ограничение на локальную плотность ПЧД с $M \sim 10^5 M_\odot$ предлагалось Ипсером и Прайсом [91]; $10^6 M_\odot$ — оценка массы облака, которое может коллапсировать в эпоху рекомбинации, и является поэтому другой возможной характерной массой ПЧД [154].

Можно, конечно, определять более сложные пределы для $\Omega_{\text{ПЧД}}$. Например, ПЧД, подобно звездам, могут образовывать гравитационно связанные системы (двойные или скопления), которые эволюционируют, излучая гравитационные волны. Если значительная часть всех ПЧД рождается в таких системах и если каждая система эволюционирует путем сжатия и слипания в единую дыру с массой $\sim M^*$ при некотором красном смещении $\sim z^*$, то в результате должно существовать изотропное гравитационное фоновое излучение с плотностью энергии $\sim 0{,}1\ \Omega_{\text{ПЧД}} \rho_c c^2/(1+z^*)$, сконцентрированное вблизи частоты $\sim c^3/[6\pi G M^* (1+z^*)]$. Если $\Omega_{\text{ПЧД}} (1+z^*) \gg$ $\gg 0{,}05\, \Omega_{\text{полн}}$, то будет виден скорее непрерывный фон, чем ряд всплесков. Гравитационно-волновые эксперименты близкого будущего могут дать интересные ограничения для таких сценариев. Конечно, если при таких событиях сравнимая мощность выделяется в электромагнитный фон [60], то этот фон будет гораздо легче обнаружить, чем гравитационно-волновой фон, при условии что он не испытывает красного смещения к частотам $\lesssim 100$ МГц.

Подводя итоги, можно сказать, что хотя представление о первичных черных дырах порождает большое количество теоретических построений и вполне может оказаться способным дать недостающее решающее звено в объединении квантовой механики и гравитации и хотя ПЧД могли бы помочь в объяснении многих загадок ранних этапов развития вселенной, пока нет наблюдательных свидетельств в пользу их существования. Вероятность обнаружить первичные черные дыры прямыми или косвенными методами по их радиационным эффектам сегодня не кажется нам значительной.

8. ЗАКЛЮЧИТЕЛЬНЫЕ ЗАМЕЧАНИЯ

Как теперь должно быть ясно, несмотря на значительные теоретические достижения последних десяти лет и взрыв новых наблюдательных данных во всем электромагнитном диапазоне, астрофизика

еще не дошла до той стадии, на которой мы можем быть твердо уверены, что черные дыры вообще существуют. Наилучшие научные доводы в пользу их существования могут быть представлены для рентгеновского источника Cygnus X-1, и здесь аргументы, которые являются достаточно неотразимыми, существенно не изменились за последние пять лет. Что касается квазаров, то теперь имеется больше свидетельств в пользу предельно компактной структуры, но это главным образом теоретические рассмотрения, ведущие к модели, описанной в разд. 6 и предусматривающей аккрецию газа $\sim 1 M_{\odot}$ в год на черную дыру с массой $\sim 10^8 M_{\odot}$. Более оптимистические надежды пятилетней давности на будущее астрофизики черных дыр, откровенно говоря, все еще не оправдались.

Тем не менее отсутствие прямого наблюдательного подтверждения не погасило энтузиазма астрономов, связанного с теоретическими изысканиями в несомненно одной из самых притягательных областей современной теоретической физики. Кроме того, хотя теоретические аргументы в пользу существования значительной плотности первичных дыр не являются достаточно убедительными, не существует серьезной альтернативы конечному гравитационному коллапсу одиночной звезды с массой $M \gtrsim 2{,}5\ M_{\odot}$ и коллапсу достаточно плотных скоплений газа и звезд; широко распространенное теоретическое предубеждение, что черные дыры действительно существуют, по-видимому, все же имеет оправдание.

Каковы в таком случае перспективы улучшить наблюдательный базис общей теории относительности, используя черные дыры, в следующие пять лет? Наибольшие надежды, по-видимому, возлагаются на рентгеновские детекторы на борту НЕАО В. Эти приборы имеют гораздо большую чувствительность и временно́е разрешение, чем их предшественники, и могли бы обнаружить быстрые пульсации с низким Q, которые, вероятно, являются наиболее надежным признаком черных дыр в квазарах и двойных рентгеновских источниках. Шансы увидеть образование черной дыры (например, через всплеск гравитационного излучения) или ее взрыв (например, через радиоимпульс) далеко не ничтожны, но здесь мы опять должны ожидать, что всякое объявленное обнаружение будет весьма проблематичным.

При этом возникает интересный вопрос. Являются ли физические условия вблизи черной дыры с неизбежностью столь беспорядочными, что общерелятивистские вычисления не способны предсказать что-либо сверх того, что установлено ньютоновским анализом, скажем, при $r=2M$? К счастью (для релятивистов, склонных к выполнению скрупулезных вычислений), уже имеется несколько примеров эффектов, отсутствующих в ньютоновском анализе, влияние которых может быть решающим при интерпретации будущих наблюдений. Два примера — это предсказанный Бардином и Петтерсоном [12] процесс увлечения аккреционного диска в экваториальную плоскость вращающейся дыры и гравитационные волны.

испускаемые в бурных событиях, протекающих с участием черных дыр. Однако, как мы пытались подчеркнуть в разд. 2, для большинства электромагнитных радиационных эффектов характерные неопределенности в микрофизике настолько значительны, что мы сомневаемся в том, что изощренные релятивистские расчеты скоро начнут играть решающую роль в теоретических моделях черных дыр и их окружения. Будем надеяться, что мы не правы.

ЛИТЕРАТУРА[1])

1. *Abell G. O.*, Exploration of the Universe, 3rd edition. Holt, Reinhart and Winston, New York, 1975.
2. *Abramowicz M.*, *Járosyński M.*, *Sikoru M.*, Astron. Astrophys., **63**, 21 (1978).
3. *Abt H. A.*, *Levy S. G.*, Astrophys. J., **188**, 291 (1974).
4. *Arnett W. D.*, Astrophys. Space Sci., **5**, 180 (1969).
5. *Arnett W. D.*, *Bowers R. L.*, Astrophys. J. Suppl. Ser., **33**, 415 (1977).
6. *Arons J.*, *Kulsrud R. M.*, *Ostriker J. P.*, Astrophys. J., **198**, 687 (1975).
7. *Auer L. H.*, *Woolf N. J.*, Astrophys. J., **142**, 182 (1965).
8. *Baade W.*, *Minkowski R.*, Astrophys. J., **119**, 206 (1954).
9. *Bahcall N. A.*, *Hausman M. A.*, Astrophys. J., **213**, 93 (1977).
10. *Bahcall J. N.*, *Ostriker J. P.*, Nature, **256**, 23 (1975).
11. *Bardeen J. M.*, *Berger B. K.*, Astrophys. J., **221**, 105 (1978).
12. *Bardeen J. M.*, *Petterson J. A.*, Astrophys. J. (Lett.), **795**, L65 (1975).
13. *Barnothy J. M.*, Astron. J., **70**, 666 (1965).
14. *Barrow J. D.*, Nature, **267**, 117 (1977).
15. *Begelman M. C.*, Mon. Not. R. Astron. Soc., **181**, 347 (1977).
16. *Begelman M. C.*, Mon. Not. R. Astron. Soc., **184**, 53 (1978).
17. *Bekenstein J. D.*, Astrophys. J., **183**, 657 (1973).
18. *Bekenstein J. D.*, Astrophys. J., **210**, 544 (1976).
19. *Berlitz C. F.*, The Bermuda Triangle. Doubleday, Garden City, New York, 1974.
20. *Bisnovatyi-Kogan G. S.*, Rev. Nuovo Cimento, **2**, ser. 3, 1 (1979).
21. *Bisnovatyi-Kogan G. S.*, *Blinnikov S. I.*, Astron. Astrophys., **59**, 111 (1977).
22. *Bisnovatyi-Kogan G. S.*, *Ruzmaikin A. A.*, Astrophys. Space Sci., **28**, 45 (1974).
23. *Bisnovatyi-Kogan G. S.*, *Ruzmaikin A. A.*, Astrophys. Space Sci., **42**, 401 (1976).
24. *Bisnovatyi-Kogan G. S.*, *Popov Yu. P.*, *Samochin A. A.*, Astrophys. Space Sci., **41**, 287 (1976).
25. *Blandford R. D.*, Mon. Not. R. Astron. Soc., **181**, 489 (1977).
26. *Blandford R. D.*, *Znajek R. L.*, Mon. Not. R. Astron. Soc., **179**, 433 (1977).
27. *Blandford R. D.*, *McKee C. F.*, *Rees M. J.*, Nature, **267**, 211 (1977).
28. *Bolton C. T.*, Nat. Phys. Sci., **240**, 124 (1972).
29. *Bolton C. T.*, Astrophys. J., **200**, 269 (1975).
30. *Bondi H.*, Mon. Not. R. Astron. Soc., **112**, 195 (1952).
31. *Buff J. et al.*, Astrophys. J., **212**, 768 (1977).
32. *Burbidge G. R.*, Comments Astrophys. Space. Sci., **4**, 105 (1972).
33. *Burbidge G. R.*, *Jones T. W.*, *O'Dell S. L.*, Astrophys. J., **193**, 43 (1974).
34. *Callahan P. S.*, Astron. Astrophys., **59**, 127 (1977).
35. *Cameron A. G. W.*, Nature, **229**, 178 (1971).
36. *Carr B. J.*, Astrophys. J., **201**, 1 (1975).
37. *Carr B. J.*, Astrophys. J., **206**, 8 (1976).
38. *Carr B. J.*, Mon. Not. R. Astron. Soc., **187**, 293 (1977).
39. *Carr B. J.*, Astron. Astrophys., **56**, 377 (1977).
40. *Carr B. J.*, *Hawking S. W.*, Mon. Not. R. Astron. Soc., **168**, 399 (1974).
41. *Carr B. J.*, *Rees M. J.*, Astron. Astrophys., **61**, 705 (1977).

[1]) Литература, отмеченная звездочкой, добавлена при переводе.— *Прим. перев.*

42. *Carter B., Gibbons G. W., Lin D. N. C., Perry M. J.*, Astron.Astrophys., **52**. 427 (1976).
43. *Chandrasekhar S.*, Astrophys. J., **74**, 81 (1931).
44. *Chandrasekhar S.*, Observatory, **57**, 373 (1934).
45. *Chapline G. F.*, Nature, **253**, 251 (1975).
46. *Christodolou D.*, Phys. Rev. Lett., **25**, 1596 (1978).
47. *Clayton D. D.*, Principles of Stellar Evolution and Nucleosynthesis, McGraw-Hill, New York, 1968.
48. *Clayton D. D., Newman M. J., Talbot Б. J. Jr*, Astrophys. J., **201**, 489 (1975).
49. *Cocke W. J., Disney M. J., Taylor D. J.*, Nature, **221**, 525 (1969).
50. *Cohen M. H. et al.*, Nature, **268**, 405 (1977).
51. *Colgate S. A.*, Astrophys. J., **150** 163 (1967).
52. *Colgate S. A., Noerdlinger P. S.*, Astrophys. J., **765**, 509 (1971).
53. *Dahlbacka G. H., Chapline G. F., Weaver T. A.*, Nature, **250**, 36 (1974).
54. *Davidson K., Ostriker J. P.*, Astrophys, J., **179**, 585 (1973).
55. *De Young D. S.*, Annu. Rev. Astron. Astrophys., **14**, 447 (1976).
56. *Detweiler S. L.*, Astrophys. J., **225**, 687 (1978).
57. *Detweiler S. L., Szedenits E., Jr.*, Astrophys. J., **231**, 211 (1979).
58. *Durisen R. H.*, Astrophys, J., **199**, 179 (1975).
59. *Eardley D. M., Press W. H.*, Annu. Rev. Astron. Astrophys., **13**, 381 (1975).
60. *Eichler D., Solinger A.*, Astrophys. J., **203**, 1 (1976).
61. *Fabian A. C., Maccagni D., Rees M. J., Stoeger W. R.*, Nature, **260**, 683 (1976).
62. *Flannery B. P., Ulrich R. K.*, Astrophys. J., **212**, 533 (1977).
63. *Forman W., Jones C., Tananbaum H.*, Astrophys. J., **208**, 849 (1976).
64. *Frank J. H., Rees M. J.*, Mon. Not. R. Astron. Soc., **176**, 633 (1976).
65. *Frautschi S.*, Phys. Rev., **D3**, 2821 (1971).
66. *Giacconi R., Gursky, H.*, X-Ray Astronomy. Reidel: Dordrecht, 1974.
67. *Gisler G. R.*, Astron. Astrophys., **51**, 137 (1976).
68. *Gnedin Yu. N., Sunyaev R. A.*, Astron. Astrophys., **25**, 233 (1973).
69. *Gold T.*, Nature, **218**, 731 (1968).
70. *Gott J. R.*, III, Astrophys. J., **173**, 227 (1972).
71. *Gott J. R.*, III, *Gunn J. E., Schramm D. N., Tinsley B. M.*, Astrophys. J., **194** 543 (1974).
72. *Greenstein J. L.*, Astron. J., **81**, 323 (1976).
73. *Greenstein J. L., Oke J. B., Shipman H. L.*, Astrophys. J., **169**, 563 (1971).
74. *Hack M. et al.*, Nature, **249**, 534 (1974).
75. *Hagedorn R.*, Nuovo Cimento Suppl., **3**, 147 (1965).
76. *Hartle J. B.*, Phys. Rep., **46**, 201 (1078).
77. *Hawking S. W.*, Mon. Not. R. Astron. Soc., **152**, 75 (1971).
78. *Hawking S. W.*, Nature, **248**, 30 (1974).
79. *Hawking S. W.*, Comm. Math. Phys., **43**, 189 (1975).
80. *Hazard C., Mitton S. A.* eds., Active Galactic Nuclei, Cambridge University Press, 1979.
81. *Hewish A. et al.*, Nature, **217**, 709 (1968).
82. *Hills J. G.*, Nature, **254**, 295 (1975).
83. *Hoyle F., Fowler W. A.*, Nature, **197**, 533 (1963).
84. *Hoyle F., Lyttleton R. A.*, Proc. Camb. Philos. Soc., **36**, 325 (1940).
85. *Hulse R. A., Taylor J. H.*, Astrophys. J. (Lett.), **195**, L51 (1975).
86. *Iben I., Jr.*, Annu. Rev. Astron. Astrophys., **12**, 215 (1974).
87. *Icke V.*, Nature, **266**, 699 (1977).
88. *Icke V.*, Astron. Astrophys., **74**, 42 (1979).
89. *Illingworth G., King I. R.*, Astrophys. J. (Lett.), **218**, L109 (1977).
90. *Ipser J. R.*, Astrophys. J., **158**, 17 (1969).
91. *Ipser J. R., Price R. H.*, Astrophys. J., **216**, 578 (1977).
92. *Jackson A. A., Ryan M. P.*, Jr., Nature, **245**, 88 (1973).
93. *Jastrow R., Thompson M. H.*, Astronomy: Fundamentals and Frontiers. Wiley, New York, 1972.
94. *Jelley J. V., Baird G. A., O'Mongain E.*, Nature, **267**, 499 (1977).
95. *Jennison R. C., Das Gupta M. K.*, Nature, **172**, 996 (1953).

96. *Kafka P., Mészáros P.*, Gen. Relativ. Grav., **7**, 841 (1976).
97. *Katz J. I.*, Astrophys. J., **208**, 910 (1976).
98. *Kellermann K. I., Pauliny-Toth I. I. K.*, Annu. Rev. Astron. Astrophys., **6**, 417 (1968).
99. *Kellermann K. I. et al.*, Astrophys. J. (Lett.), **179**, L141 (1973).
100. *Kellermann K. I. et al.*, Astrophys. J. (Lett.), **214**, L61 (1977).
101. *Kippenhahn R., Meyer-Hofmeister E.*, Astron. Astrophys., **54**, 539 (1977).
102. *Kozlowski M., Járosyński M., Abramowicz M. A.*, Astron. Astrophys., **63**, 209 (1978).
103. *Lamb F. K.*, In: Proceedings of the Eighth Texas Symposium on Relativistic Astrophysics, Ann. N. Y. Acad. Sci., **302**, 482 (1977).
104. *Lamb F. K. et al.*, Astrophys. J., **217**, 197 (1977).
105. *Lamb F. K., Pethick C. J., Pines D.*, Astrophys. J., **184**, 271 (1973).
106. *Lamers H. J. G. L. M., Morton D. C.*, Astrophys. J. Suppl. Ser., **32**, 715 (1976).
107. *Laplace P. S.*, Allgemeine geographische Ephemeriden, vol. 1, ed. F. X. von Zach, 1799, English translation in S. W. Hawking and G. F. R. Ellis (1973). The Large Scale Structure of Space-Time, appendix A. Cambridge University Press, 1973. [Имеется перевод: *Хокинг С., Эллис Дж.* Крупномасштабная структура пространства-времени.— М.: Мир, 1976, с. 406.]
108. *Liang E. P. T., Price R. H.*, Astrophys. J., **218**, 247 (1977).
109. *Lightman A. P.*, Astrophys. J., **194**, 419 (1974).
110. *Lightman A. P.*, Astrophys. J., **194**, 429 (1974).
111. *Lightman A. P., Eardley D. M.*, Astrophys. J. (Lett.), **187**, L1 (1974).
112. *Lightman A. P., Fall M.*, Astrophys. J., **221**, 567 (1978).
113. *Lightman A. P., Shapiro S. L.*, Astrophys. J., **211**, 244 (1977).
114. *Lightman A. P., Shapiro S. L.*, Rev. Mod. Phys., **50**, 437 (1978).
115. *Lightman A. P., Press W. H., Odenwald S. F.*, Astrophys. J., **219**, 629 (1978).
116. *Lin D. N. C., Carr B. J., Fall S. M.*, Mon. Not. R. Astron. Soc., **177**, 51 (1976).
117. *Lynden-Bell D.*, Nature, **223**, 690 (1969).
118. *Lynden-Bell D., Pringle J. E.*, Mon. Not. R. Astron. Soc., **168**, 603 (1974).
119. *Lynden-Bell D., Rees M. J.*, Mon. Not. R. Astron. Soc., **152**, 461 (1971).
120. *McCray R., Hatchett S. P.*, Astrophys. J., **199**, 196 (1975).
121. *Manchester R. N., Taylor J. H.*, Pulsars. W. H., Freeman, San Francisco, 1977. [Имеется перевод: *Манчестер Р. Н., Тэйлор Дж. Х.* Пульсары.— М.: Мир, 1980.]
122. *May M. M., White R. H.*, Phys. Rev., **141**, 1232 (1966).
123. *Mazurek T. J., Meier D. L., Wheeler J. C.*, Astrophys. J., **213**, 518 (1977).
124. *Meier D. L., et al.*, Astrophys. J., **204**, 869 (1976).
125. *Meikle W. P. S.*, Nature, **269**, 41 (1977).
126. *Menzel D. H., Whippe F. L., De Vaucouleurs G.*, Survey of the Universe. Prentice-Hall, Englewood Cliffs, New Jersey, 1970.
127. *Mészáros P.*, Astron. Astrophys., **37**, 225 (1974).
128. *Mészáros P.*, Astron. Astrophys., **38**, 5 (1975).
129. *Mészáros P.*, Astron. Astrophys., **44**, 59 (1975).
130. *Mészáros P.*, Nature, **258**, 583 (1975).
131. *Michel F. C.*, Astrophys. Space Sci., **15**, 153 (1972).
132. *Misner C. W.*, Battelle Rencontres in Mathematics and Physics, eds. C. DeWitt, B. DeWitt, Benjamin Press, New York, 1968.
133. *Misner C. W., Thorne K. S., Wheeler J. A.*, Gravitation. W. H. Freeman, San Francisco, 1973. [Имеется перевод: *Мизнер Ч. В., Торн К. С., Уилер Дж. А.* Гравитация.— М.: Мир, 1977.]
134. *Morrison P.*, Astrophys. J. (Lett.), **157**, L73 (1969).
135. *Novikov I. D., Thorne K. S.*, In: Black Holes, eds. C. DeWitt, B. DeWitt, p. 343. Gordon and Breach, New York, 1973.
136. *Novikov I. D., Zel'dovich Ya. B.*, Nuovo Cimento Suppl., **4**, 810, addendum 2 (1966).
137. *Oppenheimer J. R., Snyder H.*, Phys. Rev., **56**, 455 (1939). [Имеется перевод: Альберт Эйнштейн и теория гравитации.— М.: Мир, 1979, с. 353.]

138. *Oppenheimer J. R., Volkoff G.*, Phys. Rev., **55**, 374 (1939). [Имеется перевод: Альберт Эйнштейн и теория гравитации.— М.: Мир, 1979, с. 337.]
139. *Osaki Y.*, Publ. Astron. Soc. Jap., **26**, 429 (1974).
140. *Ostriker J. P.*, In: Proceedings of the Eighth Texas Symposium on Relativistic Astrophysics, Ann. N. Y. Acad. Sci., **302**, 229 (1977).
141. *Ostriker J. P., Peebles P. J. E.*, Astrophys. J., **186**, 467 (1973).
142. *Ostriker J. P., Richstone D. O., Thuan T. X.*, Astrophys, J. (Lett.), **188**, L87 (1974).
143. *Озерной Л. М.*, Астрон. журн., **43**, 300 (1966).
144. *Озерной Л. М., Чертопруд В. Е.*, Астрон. журн., **46**, 940 (1969).
145. *Ozernoy L. M., Usov V. V.*, Astrophys. Space Sci., **25**, 149 (1973).
146. *Pacini F.*, Nature, **216**, 567 (1967).
147. *Paczynski B.*, Acta Astron., **28**, 91 (1978); **28**, 241 (1978).
148. *Page D. N.*, Phys. Rev., **D13**, 198 (1976).
149. *Page D. N.*, Phys. Rev., **D14**, 3260 (1976).
150. *Page D. N.*, Phys. Rev., **D16**, 2402 (1977).
151. *Page D. N., Hawking, S. W.*, Astrophys. J., **206**, 1 (1976).
152. *Payne-Gaposchkin C., Haramundanis K.*, Introduction to Astronomy. Prentice-Hall, Englewood Cliffs, New Jersey, 1970.
153. *Peebles P. J. E.*, Physical Cosmology, Princeton University Press, New Jersey, 1971. [Имеется перевод: *Пиблс П. Дж. Е.* Физическая космология.— М.: Мир, 1975.]
154. *Peebles P. J. E., Dicke R. H.*, Astrophys. J., **154**, 891 (1968).
155. *Penrose R.*, Nuovo Cimento **1**, special number, 252, специальный выпуск, 1969.
156. *Piddington J. N.*, Mon. Not. R. Astron. Soc., **148**, 131 (1970).
157. *Piran T.*, Mon. Not. R. Astron. Soc., **180**, **45** (1977).
158. *Pratt J. P., Strittmatter P. A.*, Astrophys. J. (Lett.), **204**, L29 (1976).
159. *Press W. H., Gunn J. E.*, Astrophys. J., **185**, 397 (1973).
160. *Pringle J. E.*, Mon. Not. R. Astron. Soc., **177**, 65 (1976).
161. *Pringle J. E.*, In: Proceedings of the Eighth Texas Symposium on Relativistic Astrophysics, Ann. N. Y. Acad. Sci., **302**, 6 (1977).
162. *Pringle J. E., Rees M. J.*, Astron. Astrophys, **21**, 1 (1972).
163. *Pringle J. E., Rees M. J., Pacholczyk A. G.*, Astron. Astrophys., **29**, 179 (1973).
164. *Rappaport S. A., Joss P. C.*, In: Proceedings of the Eighth Texas Symposium on Relativistic Astrophysics, Ann. N. Y. Acad. Sci., **302**, 460 (1977).
165. *Readhead A. C. S., Cohen M. H., Blandford R. D.*, Nature, **272**, 131 (1978).
166. *Rees M. J.*, Phys. Rev. Lett., **28**, 1660 (1972).
167. *Rees M. J.*, Nature, **266**, 333 (1977).
168. *Rees M. J.*, In: Proceedings of the Eighth Texas Symposium on Relativistic Astrophysics, Ann. N. Y. Acad. Sci., **302**, 613 (1977).
169. *Refsdal S.*, Mon. Not. R. Astron. Soc., **128**, 295 (1964).
170. *Robinson I., Schild A., Schucking E. L.*, Quasi-Stellar Sources and Gravitational Collapse. University of Chicago Press, Chicago, 1965.
171. *Rothschild R. E. et al.*, Astrophys. J. (Lett.), **189**, L13 (1974).
172. *Rothschild R. E. et al.*, Astrophys. J. (Lett.), **213**, 818 (1977).
173. *Ryan M. P., Jr.*, Astrophys. J. (Lett.), **177**, L79 (1972).
174. *Salpeter E. E.*, Astrophys. J., **140**, 796 (1964).
175. *Sargent W. L. W. et al.*, Astrophys. J., **221**, 731 (1978).
176. *Savonije G. J.*, Astron. Astrophys., **62**, 317 (1978).
177. *Schmidt M.*, Nature, **197**, 1040 (1963).
178. *Shakura N. I., Sunyaev R. A.*, Mon. Not. R. Astron. Soc., **175**, 613 (1976).
179. *Shapiro S. L.*, Astrophys. J., **180**, 531 (1973).
180. *Shapiro S. L.*, Astrophys. J., **185**, 69 (1973).
181. *Shapiro S. L., Lightman A. P.*, Astrophys. J., **204**, 555 (1976).
182. *Shapiro S. L., Lightman A. P., Eardley D. M.*, Astrophys. J., **204**, 187 (1976).
183. *Shields G. A., Wheeler J. C.*, Astrophys. Lett., **17**, 69 (1976).
184. *Shklovsky I. S.*, Astrophys. J. (Lett.), **148**, L1 (1967).

185. *Шварцман В. Ф.*, Астрон. журн., **48**, 479 (1971).
186. *Silk J.*, *Arons J.*, Astrophys. J. (Lett.), **200**, L131 (1975).
187. *Smak J.*, New Directions and New Frontiers in Variable Star Research, IAU Colloquium No. 15, Veröffentlichung der Remeis—Sternwarte, Bamberg, **9**, 248 (1971).
188. *Smak J.*, Acta Astron., **26**, 277 (1976).
189. *Smarr L.*, In: Proceedings of the Eighth Texas Symposium on Relativistic Astrophysics, Ann. N. Y. Acad. Sci., **302**, 569 (1977).
190. *Spitzer L. Jr.*, Physics of Fully Ionized Gases, 2nd edition, Interscience, New York, 1962. [Имеется перевод: *Спитцер Л.* Физика полностью ионизованного газа.— М.: Мир, 1965.]
191. *Spitzer L. Jr.*, *Shull J. M.*, Astrophys. J., **201**, 773 (1975).
192. *Spitzer L. Jr. Thuan T. X.*, Astrophys. J., **175**, 31 (1972).
193. *Stothers R.*, *Ezer D.*, Astrophys. Lett., **13**, 45 (1973).
194. *Sturrock P. A.*, *Barnes, C.*, Astrophys. J., **176**, 31 (1972).
195. *Сюняев Р. А.*, Астрон. журн., **49**, 1153 (1972).
196. *Taam R. E.*, *Bodenheimer P.*, *Ostriker J. P.*, Astrophys. J., **222**, 269 (1978).
197. *Tammann G. A.*, In: Proceedings of the Eighth Texas Symposium on Relativistic Astrophysics, Ann. N. Y. Acad. Sci., **302**, 61 (1977).
198. *Taylor J. H.*, *Manchester R. N.*, Annu. Rev. Astron. Astrophys., **15**, 19 (1977).
199. *Thomas H.-C.*, Annu. Rev. Astron. Astrophys., **15**, 127 (1977).
200. *Thorne K. S.*, In: Theoretical Principles in Astrophysics and Relativity, eds. N. Lebovitz, W. Reid and P. Vandervoort. University of Chicago Press, Chicago, 1978.
201. *Thorne K. S.*, *Żytkow A.*, Astrophys. J., **212**, 832 (1977).
202. *Thorstensen J. R.*, *Partridge R. B.*, Astrophys. J., **200**, 527 (1975).
203. *Thuan T. X.*, *Oke J. B.*, Astrophys. J., **205**, 360 (1976).
204. *Tinsley B. M.*, *Larsen R. B.*, eds., the Evolution of Galaxies and Stellar Populations. Yale University Observatory, New Haven, 1977.
205. *Tremaine S. D.*, Astrophys. J., **203**, 345 (1976).
206. *Trimble V.*, *Greenstein J. L.*, Astrophys. J., **177**, 441 (1972).
207. *Trimble V.*, *Thorne K. S.*, Astrophys. J., **156**, 1013 (1969).
208. *Вайнер Б. В.*, *Нассельский П. Д.*, Письма в Астрон. журн., **3**, 147 (1977).
209. *Van den Bergh S*, Nature, **224**, 891 (1969).
210. *Van den Bergh S.*, Annu. Rev. Astron. Astrophys., **13**, 217 (1975).
211. *Van den Heuvel E. P. J.*, Astrophys. J. (Lett.), **198**, L109 (1975).
212. *Van den Heuvel E. P. J.*, *Heise J.*, Nature, Phys. Sci., **239**, 67 (1972).
213. *Van Riper K.*, Astrophys. J., **221**, 304 (1978).
214. *Véron M. P.*, Astron. Astrophys., **41**, 423 (1975).
215. *Webster B. L.*, *Murdin P.*, Nature, **235**, 37 (1972).
216. *Weidemann V.*, Annu. Rev. Astron. Astrophys., **6**, 351 (1968).
217. *Weidemann V.*, Astron. Astrophys., **59**, 411 (1977).
218. *Weisskopf M. C.*, *Sutherland P. G.*, Astrophys. J., **221**, 228 (1978).
219. *Wheeler J. A.*, In: La structure et l'evolution de l'univers, chapter by B. K. Harrison, M. Wakano and J. A. Wheeler, p. 124. Editions Stoops, Brussels, 1958.
220. *Wheeler J. A.*, In: Relativity, Groups, and Topology, ed. C. M. DeWitt and P. S. DeWitt, p. 317. Gordon and Breach, New York, 1964.
221. *Wheeler J. A.*, Am. Sci., **56**, 1 (1968).
222. *Whelan J. A. J. et al.*, Mon. Not. R. Astron. Soc., **181**, 259 (1977).
223. *Willis A. G.*, *Strom R. G.*, *Wilson A. S.*, Nature, **250**, 625 (1974).
224. *Wollman E. R. et al.*, Astrophys. J. (Lett.), **218**, L103 (1977).
225. *Young P. J.*, Astrophys. J., **215**, 36 (1977).
226. *Young P. J. et al.*, Astrophys. J., **221**, 721 (1978).
227. *Зельдович Я. Б.*, ЖЭТФ, **41**, 1609 (1961).
228. *Зельдович Я. Б.*, ДАН СССР, **155**, 67 (1964).
229. *Zel'dovich Ya. B.*, *Guseynov O. H.*, Astrophys. J., **144**, 840 (1966).
230. *Зельдович Я. Б.*, *Новиков И. Д.* Релятивистская астрофизика.— М.: Наука, 1967.

231. *Зельдович Я. Б., Подурец М. А.*, Астрон. журн., **42**, 963 (1965).
232. *Зельдович Я. Б., Старобинский А. А.*, Письма в ЖЭТФ, **24**, 616 (1976).
233. *Зельдович Я. Б. и др.*, Письма в Астрон. журн., **3**, 208 (1977).
234*. *Бакулин П. И., Кононович Э. В., Мороз В. И.* Курс общей астрономии.— М.: Наука, 1977.
235*. *Гинзбург В. Л.* Теоретическая физика и астрофизика.— М.: Наука, 1981.
236*. *Соболев В. В.* Курс теоретической астрофизики.— М.: Наука, 1967.
237*. *Мартынов Д. Я.* Курс общей астрофизики.— М.: Наука, 1979.
238*. *Мартынов Д. Я.* Курс практической астрофизики.— М.: Наука, 1977.
239*. *Зельдович Я. Б., Новиков И. Д.* Теория тяготения и эволюция звезд.— М.: Наука, 1971.

IV. КОСМОЛОГИЯ И РАННЯЯ ВСЕЛЕННАЯ

Я. Б. Зельдович

1. ВВЕДЕНИЕ

В книге, посвященной столетию со дня рождения Эйнштейна, уместно начать эту главу с некоторых замечаний, касающихся его личного влияния на космологию. Уже в 1917 г., непосредственно после открытия общей теории относительности, Эйнштейн написал работу, озаглавленную «Kosmologische Betrachtungen zur allgemeinen Relativitätstheorie» [1]) [17].

В настоящее время мы уже знаем из опыта, что каждое значительное продвижение по шкале исследуемых масштабов ведет к новым принципам. Так это было, когда в масштабе атома приобрели первостепенное значение квантовые свойства вещества. Возможно, что теперь, в 70-е годы, к этому можно было бы добавить теорию невылетания кварков, действующую на субъядерном уровне.

Общая теория относительности была задумана как теория, необходимая для больших, может быть бесконечных, масштабов вселенной как целого. Поэтому ее применение к космологии находилось в полном соответствии с внутренней логикой науки. К тому же не требовалось никакой другой теории, специально разработанной для больших масштабов,— это было бы излишним, поскольку общая теория относительности не содержала внутренних противоречий и расхождений с экспериментом. Первоначальная статья Эйнштейна по космологии была не безупречна, но, как утверждал Бор, даже ошибки гения представляют интерес, и их исправление ведет к важным открытиям.

Эйнштейн предположил *a priori* неоднородность вселенной в больших масштабах; это мнение теперь подтверждается в масштабах, больших 1000 Мпс (сравните с $c/H=6000$ Мпс — «горизонтом», т. е. предельной глубиной проникновения любого мыслимого наблюдения).

Но он также предположил, что вселенная статична и не эволюционирует, что оказалось неправильным. В действительности вселенная является расширяющейся системой с определенным сценарием эволюции. Противоречие с реальной ситуацией привело также Эйнштейна к теоретическим трудностям: он пришел к выводу, что ньютоновская теория гравитации неприменима к космологии, и

---

[1]) Космологические приложения к общей теории относительности (*нем.*).— *Прим. перев.*

изменил уравнения общей теории относительности, введя «космологический член». Теперь хорошо известно, что трудности ньютоновской теории (например, «гравитационный парадокс») являются следствием предубеждения. Точный последовательный анализ показывает, что ньютоновский подход к космологии расширяющейся вселенной является вполне корректным; требуется только моральная смелость. В простом виде такая теория была построена Милном [43]. Обзор с усовершенствованием понятия однородности дан автором [73].

По иронии судьбы простой ньютоновский подход был сформулирован через десять лет после того, как были найдены гораздо более сложные общерелятивистские решения Фридмана [19, 20]. Ньютоновское доказательство было найдено тогда, когда ответ был уже известен из релятивистской теории, созданной Эйнштейном.

Один вопрос, не решенный еще и теперь, касается плотности вещества во вселенной. Кроме того, вопрос о том, является ли вселенная закрытой или открытой, не решен окончательно. Но даже довольно широкие ограничения, даваемые наблюдательными данными, приводят к информации об элементарных частицах, которая не может быть получена обычными методами лабораторной и ускорительной физики. Этот вопрос рассматривается в разд. 2.

Принятие эволюционной теории, включающей динамику расширения вселенной, дает первое приближение к структуре вселенной в больших масштабах: геометрический, кинематический и динамический фон. Следующий шаг состоит в выяснении физического содержания, присутствующих частиц и температуры среды.

Было сделано несколько попыток объяснить изобилие различных химических элементов и изотопов нуклеосинтезом в однородно расширяющемся веществе. Они были основаны на завышенных оценках постоянной Хаббла H, которую сначала считали равной 550 км$\cdot$с$^{-1}\cdot$Мпс$^{-1}$.

Полученный в результате возраст $H^{-1}=2\cdot 10^9$ лет был меньше или равен оценкам возраста Земли, Солнца и шаровых скоплений. Поэтому представление о космологическом возникновении всех элементов казалось убедительным.

Гамов [21] пытался сделать это, рассматривая ванну, полную нейтронов, в которой происходят поглощение нейтронов ядрами и последующие бета-распады. Чтобы помешать изотопам сильно реагировать с медленными нейтронами, Гамов [22, 23] ввел высокие начальные температуры и вычислил современное значение температуры, равное 6 K, что поразительно близко к действительно измеренной температуре 2,7 K. Все промежуточные стадии расчета Гамова оказались несостоятельными: возраст вселенной равен $(15-20)\times \times 10^9$ лет — гораздо больше, чем возраст Земли и Солнца ($\sim 5\cdot 10^9$) лет. Космологический нуклеосинтез практически заканчивается на He^4; элементы со средними и большими атомными весами образуются в звездах.

Однако горячая модель «большого взрыва», предложенная Гамовым, была развита дальше в работах Альфера и Германа [1], Хайяши [30], Ферми и Туркевича (цитируется в работе [1]) и др. В работе, которая осталась незамеченной радиоастрономами, Дорошкевич и Новиков [16] указали, что даже при низкой температуре космологическое излучение должно превосходить излучение звезд, радиоисточников и т. д. в определенном диапазоне длин волн (например, для λ между 0,1 и 20 см).

Хорошо известно, что Дикки и др. [13] независимо пришли к тому же выводу и, более того, сделали попытку найти космологическое излучение с длиной волны 3 см. Поэтому, когда Пензиас и Вильсон [54] действительно обнаружили шум на 7,3 см, Дикки и его сотрудники сразу поняли его происхождение и объявили, что теория «большого взрыва» доказана. Позднее было осознано, что наблюдавшееся ранее Мак-Келларом [42] необъяснимое возбуждение межзвездных молекул CN также было доказательством и к тому же давало хорошую меру космологического излучения. Горячая модель «большого взрыва» приводит к естественному делению космологического сценария. В порядке возрастающей отдаленности от нашего времени, а также возрастающей сложности рассмотрим следующие этапы.

1. Лептонная эра: от $t=0{,}01$ с, $z=10^{11}$, $T=3\cdot10^{11}$ К=30 МэВ вплоть до ситуации на сегодняшний день. Физические законы в эту эпоху хорошо известны, космология помогает установить важные детали. Лептонная эра рассматривается в разд. 3.

2. Адронная эра: от 10^{-42} до 0,01 с. Развитие физики частиц, включая нарушение симметрий и кварковый конфайнмент, демонстрирует возможность многих неожиданных новых качественных свойств горячей плазмы в этой области (разд. 4).

3. Последнее, но не менее важное. Планковская единица времени 10^{-43} с как раз характеризуется прямым воздействием квантовой теории на космологию, гравитацию и метрику. Фактически остальные статьи в этой книге посвящены этому предмету. Испарение черных дыр — квантовый эффект, который продолжается за планковское время. Некоторые замечания о первичных черных дырах и квантовой гравитации даны в разд. 5.

2. СРЕДНЯЯ ПЛОТНОСТЬ ВЕЩЕСТВА ВО ВСЕЛЕННОЙ

Плотность вещества, непосредственно видимого как звезды и светящиеся газовые облака в галактиках, усредненная по всему объему, включающему темное межгалактическое пространство, является величиной порядка 10^{-30}—10^{-31} г/см^3.

Плотность энергии электромагнитного излучения космологического происхождения (чернотельное реликтовое излучение с температурой 2,7 К) равна $5\cdot10^{-13}$ эрг/см^3. Эта величина, деленная на c^2, дает значение плотности $5\cdot10^{-34}$ г/см^3, которое в настоящее вре-

мя пренебрежимо мало по сравнению с плотностью обычного вещества.

Физикам теперь известно о существовании слабо взаимодействующих частиц — в первую очередь нейтрино. Эта ситуация иллюстрируется поиском солнечных нейтрино: считается, что их поток составляет 5% от солнечного светового потока, т. е. 10^5 эрг/(см$^2\cdot$с), что соответствует плотности энергии $3\cdot10^{-6}$ эрг/см^3 и плотности массы $3\cdot10^{-27}$ г/см^3, однако они еще не обнаружены. Электронные нейтрино с более низкой энергией на одну частицу и другие типы нейтрино, а также гравитоны обнаружить гораздо труднее по сравнению с солнечными мегавольтными нейтрино.

Это значит, что прямыми физическими измерениями в лаборатории невозможно обнаружить нейтрино даже в тех количествах, которые уже чрезвычайно важны в космологии. В космологическом масштабе гравитация оказывается более чувствительной к нейтрино, чем лабораторные детекторы частиц [77].

Первым и очевидным результатом избытка плотности было бы уменьшение времени эволюции:

$$t=\frac{1}{H}f(\Omega),\qquad \Omega=\frac{\rho 8\pi G}{3H^2}=\frac{\rho}{\rho_c},\quad f(0)=1,$$

$f<1$ для $\Omega>0$; $f=(1+\sqrt{\Omega})^{-1}$ для релятивистских частиц. Для $\Omega\gg1$ t асимптотически не зависит от H:

$$t\leqslant(3/8\pi G\rho)^{1/2}.$$

Принимая возраст солнечной системы $4{,}5\cdot10^9$ лет за нижнюю границу для t, получаем $\rho<10^{-28}$ г/см^3. Но это очень завышенная оценка. С большой уверенностью можно положить $\rho<10^{-29}$ г/см^3.

Это позволяет дать верхние пределы для температуры равновесного ферми-газа ν—$\bar\nu$ или чернотельного гравитационного излучения: $T<40$ К.

Другая возможность — море вырожденных нейтрино (с пренебрежимо малой примесью $\bar\nu$ или, наоборот, вырожденные $\bar\nu$ без ν). Она согласуется со сценарием горячей модели «большого взрыва». Верхний предел для ферми-энергии этого моря порядка 50 эВ.

Это все, что можно сказать о количествах частиц. Космология дает также информацию о свойствах нейтрино, т. е. о их массе покоя. Предположение, что нейтрино имеет ненулевую массу покоя, является произвольным, но его трудно опровергнуть средствами ядерной физики. Экспериментально установлено, что $m<100$ эВ для электронных нейтрино, $m<5\cdot10^5$ эВ для мюонных нейтрино и $m<5\cdot10^8$ эВ для нейтрино, связанных с новым «лептоном» τ — тяжелым ($m=1{,}85\cdot10^9$ эВ) аналогом электрона и мюона. Космологические ограничения впервые использовали Герштейн и Зельдович [24]. В горячей модели «большого взрыва» плотность числа различных нейтрино должна быть порядка 200 см$^{-3}$, если их масса покоя

меньше 10^6 эВ, так что они не аннигилируют в ходе расширения.

Используя завышенную оценку $\rho<10^{-28}$ г/см^3, они получили для масс покоя ν_e или ν_μ приблизительный верхний предел $m<400$ эВ. Это было не важно для ν_e, но очень важно для мюонных нейтрино.

Каусик и Мак-Клелланд [12] усовершенствовали теорию, приняв $\rho\leqslant5\cdot10^{-30}$ г/см^3, а также равную массу (если она есть) для всех типов нейтрино, и получили более сильный результат $m<13$ эВ. Салаи и Маркс [64] рассмотрели судьбу тяжелых нейтрино в галактиках.

После открытия третьего лептона τ много групп исследователей независимо использовали космологические аргументы, чтобы исключить интервал 10^6 эВ$>m>$20 эВ возможных значений масс покоя ν_τ. Однако интервал $6\cdot10^8$ эВ $>m>10^6$ эВ ставит новые проблемы: аннигиляция ν_τ и $\bar\nu_\tau$ в ходе расширения имеет место, когда температура падает до $kT<m_\nu c^2$. Была использована теория аннигиляции в ходе расширения, развитая в связи с аннигиляцией вещество — антивещество [10] и кварковой аннигиляцией [81]. Ее важным результатом является алгебраическая (не экспоненциальная) зависимость окончательной концентрации от массы покоя.

В результате [37, 38, 67] появился запрет на интервал масс покоя, приведенный выше. Поэтому масса покоя третьего нейтрино также должна быть меньше 20 эВ, и точно нулевое значение является весьма вероятным. Время жизни массивных нейтрино обсуждалось в работе Дикуса, Кольба и Теплица [14]. Случай кварков, упомянутый выше, был важен для развития теории элементарных частиц. Если свободные кварки могут существовать, их концентрация после рекомбинации должна быть равна $(Gm^2/\hbar c)^{1/2}\sim10^{-17}$ концентрации фотонов. Отношение числа кварков к числу барионов должно быть порядка 10^{-8}. Отсутствие такой концентрации в обычной материи было аргументом в пользу теории невылетания кварков. Этот аргумент был сильнее, чем результаты экспериментов на ускорителях и с космическими лучами. Возможные свойства хиггсовского мезона проанализированы с использованием космологических аргументов в работе [57].

3. ЛЕПТОННАЯ ЭРА

Лептонная фаза, начинающаяся, скажем, от 0,1 или 1 с, дает важные ключи к волнующим загадкам физики и космологии.

В стандартном сценарии горячей модели равновесное нейтрон-протонное превращение замораживается при 15% n, 85% p, после чего путем нуклеосинтеза получается первичный состав 30% He^4, 70% H. С другой стороны, позитроны аннигилируют с электронами, оставляя только избыток e^-, необходимый для компенсации заряда α-частиц и протонов. Равновесный планковский спектр фотонов

поддерживается поглощением, излучением и рассеянием. После этого следует длинный период, когда равновесие сохраняется в силу того, что расширение уменьшает температуру, не возмущая форму электромагнитного спектра.

При минимальном значении температуры $T \sim 4000$ К происходит отделение материи (которая теперь нейтральна) от излучения. Гравитационная неустойчивость ведет к образованию галактик из малых начальных возмущений плотности. Каждая фаза этого процесса чувствительна к различным возмущениям стандартного сценария. Шварцман [60] указал, что взрыв заселенности среди безмассовых или легких нейтральных частиц (например, двадцати типов нейтрино) должен был бы изменить темп изменения температуры, поэтому изменилось бы отношение n : p и возросло содержание He^4 в первичном веществе. Штейгман и др. [63] утверждают, что число типов безмассовых частиц меньше семи.

Можно ожидать, что частицы, распадающиеся или аннигилирующие позже момента $t=10^4$ с (красное смещение z меньше 10^7), будут перекачивать энергию в излучение и портить планковский спектр. То же самое будет происходить с распадающимися акустическими колебаниями, если начальные возмущения соответствующего масштаба будут больше 10% плотности.

Чувствительность спектра к потреблению энергии была исследована Зельдовичем, Илларионовым и Сюняевым [82]. Благодаря переформированию спектра фотонов комптоновским рассеянием потребление 1% энергии ведет к увеличению среднего числа фотонов в наиболее заселенной части спектра примерно на 1%, но имеется чувствительная спектральная область (в рэлей-джинсовской длинноволновой части при длине волны ~ 20 см), где плотность фотонов и эффективная температура после подвода энергии понижаются на 10—20%. Эти эффекты пока не обнаружены, но благодаря высокой чувствительности даже их отсутствие дает ценные верхние пределы для отдельных типов возмущений, накладываемых на идеализированный сценарий.

4. АДРОННАЯ ЭРА

При высоких температурах адронной эры теория предсказывает изобилие барионов и антибарионов — «материя» и «антиматерия» в термодинамическом равновесии. На самом деле в лептонную эпоху имеется малая примесь (10^{-8}—10^{-9}) барионов к фотонам и нейтрино. Будучи экстраполировано к адронной эре, это означает, что имеется слабая асимметрия $(1+10^{-8})$: 1 отношения барионы/антибарионы в прошлом. Последнее кажется очень странным.

Аргументы, связанные с образованием галактик и спектром фонового излучения (2,7 К), приводят к выводам относительно пространственного распределения этого отношения (экстраполированного к адронной эре). В большом масштабе (мегапарсеки сопут-

ствующих координат[1])) вариации меньше 1% избытка, от $1+1{,}01\times\times10^{-8}$ до $1+0{,}99\cdot10^{-8}$.

В малом масштабе, но большем чем 1 пс возможны бо́льшие вариации, но тем не менее это отношение никогда не переходит в обратное ($r>1$), так что областей с избытком антивещества не существует.

Существование таких областей в масштабе, даже меньшем чем 1 пс, невозможно опровергнуть при условии, что аннигиляция с барионным избытком в среде кончается в течение адронной эры, но до периода нуклеосинтеза и спектрально-чувствительного периода лептонной эры.

Термодинамические флуктуации с $\Delta N/N=1/\sqrt{N}$ должны давать $r<1$ в областях с $N\approx\bar{N}=10^{16}$, $|N-\bar{N}|\sim10^8$, что соответствует сопутствующему масштабу порядка одного километра. Эти мелкомасштабные флуктуации должны распадаться в результате диффузии задолго до лептонной эры.

Приведенная выше информация получена из наблюдений. Насколько известно автору, нет никакой твердой фундаментальной теории, объясняющей наблюдаемое значение r, и не существует никаких теоретических утверждений относительно его пространственной однородности или изменения.

Вторая проблема, касающаяся адронной эры, связана с теориями нарушения симметрии. Предполагается, что одно или несколько волновых полей Ψ характеризуется специфической зависимостью потенциальной энергии от амплитуды $V(\Psi)$ типа $A-B\Psi^2+C\Psi^4$, где A, B, $C>0$. Эта функция имеет минимумы при $\Psi=\pm(B/2C)^{1/2}==\pm\Psi_m$.

Холодному вакууму соответствует $\Psi=\pm\Psi_m\neq0$. Например, Т. Д. Ли [39] использует существование двух решений, чтобы построить теорию, которая симметрична в своих начальных предположениях (включая симметрию $\Psi\rightarrow-\Psi$). Но в заданной области пространства, например с $\Psi=-\Psi_m$, симметрия нарушается, свойства барионов и антибарионов различны (так называемое нарушение CP-инвариантности и несохранение четности). В других теориях массы частиц зависят от Ψ_m, являясь отличными от нуля.

Киржниц [31] указал (и его коллега Линде в дальнейшем детально разработал [32—34, 40]), что при достаточно высокой температуре симметрия вакуума восстанавливается, $\bar{\Psi}=0$ везде. При уменьшении температуры происходит фазовый переход от симметричной фазы с $\Psi=0$ к антисимметричной фазе $|\bar{\Psi}|=\Psi_m$. Однако знаки $\bar{\Psi}$ в удаленных областях (вне горизонта) не коррелированы. Поэтому должны встречаться граничные слои — «стенки», разделяющие $\bar{\Psi}=+\Psi_m$ и $\bar{\Psi}=-\Psi_m$. Космологические следствия, включая массу «стенок» и связанные с ними возмущения, были рассмотрены

[1]) Измеренные в современной шкале расстояний после расширения.

Зельдовичем, Кобзаревым и Окунем [83]. Результат оказывается отрицательным для этого типа теории. Другие варианты с нестабильным вакуумом в одном состоянии, спонтанно распадающимся в другое более низкое состояние, рассматриваются Коулменом [11], Фрэмптоном [18], Кобзаревым, Окунем и Волошиным [35]. Ситуация меняется, если Ψ комплексно и вместо Ψ^2 в $V(\Psi)$ записывается $|\Psi|^2=\Psi\overline{\Psi}$.

Холодный вакуум опять не симметричен, $|\overline{\Psi}|=\Psi_m$, а горячий вакуум симметричен $\Psi\equiv 0$. Но вырождение холодного вакуума теперь является непрерывным: всякое состояние

$$\Psi=\Psi_m e^{i\varphi}$$

с произвольной фазой φ является кандидатом в холодный вакуум. Фазовый переход из горячей $\Psi=0$ в холодную фазу оставляет вихревые линии (линии, вдоль которых фаза меняется на 2π) вместо стенок.

Детальная теория вихревого поведения до сих пор не развита. В этом виновата не леность космологов. Ясно, что сначала физики должны сделать свой выбор типа теории нарушения симметрии (если она есть), получить лабораторное подтверждение этой теории и найти параметры.

Мораль вышеизложенного заключается в большом числе качественно различных вариантов. Примерно десять или двадцать лет назад многие (включая автора этой главы) думали, что прогресс физики частиц может внести лишь количественные изменения в уравнение состояния (давление как функция плотности и энтропии). Из-за быстрого достижения равновесия это означало бы, что лептонная эра не зависит от неизвестных деталей адронной эры (за исключением отношения лептонов и барионов к значениям удельной энтропии). Фазовые переходы интересны тем, что в принципе они должны быть источниками возмущений, приводящих позднее к образованию галактик в космологической модели с идеальной начальной однородностью.

Первые идеи, касающиеся фазовых переходов, были сформулированы Омнесом [46—48]. Он высказал гипотезу, что имеется интервал температур $0{,}3\ m_pc^2<kT<nm_pc^2$ (с неизвестным $n>1$), где однородная зарядово-симметричная смесь барионов и антибарионов (а также лептонов и мезонов) в термодинамическом равновесии нестабильна. Согласно Омнесу, барионы и антибарионы отталкивают друг друга, и формируются две стабильные фазы — с избытком вещества и с избытком антивещества.

Позднейшие рассмотрения заронили сомнение относительно этого частного типа фазового перехода (ядерная физика [3]; космология, «Факты против антивещества» [61, 62]). Будущее теорий нарушения симметрии также не ясно: либо они являются только математическим инструментом для перенормировки теории и позднее

исчезнут, либо истина найдена и фазовые переходы типа переходов Киржница существуют реально.

Неисследованными остаются также уравнения состояния и качественные особенности плотной горячей равновесной смеси кварков, антикварков и глюонов с учетом специфических свойств глюонов (см. [8] — кварки при высоких плотностях).

Известно или по крайней мере предполагается, что свободный цветовой заряд кварков ведет к дальнодействующим глюонным силам, даже более сильным, чем электростатические силы. Поэтому в равновесии крупномасштабные флуктуации цвета подавляются по сравнению с флуктуациями барионного заряда. Но мы не имеем никакой теории начальной ситуации в сингулярности. Поэтому можно было бы спросить: что должно произойти, если крупномасштабные флуктуации цвета заданы в первоначальном сингулярном состоянии? Тот же самый вопрос мог бы быть задан относительно электрического заряда и магнитного поля (эквивалентного токам) в большом масштабе. Ответ на эти вопросы неизвестен, и мы еще раз видим, насколько богатым является выбор априорных космологических моделей.

5. КВАНТОВАЯ ЭРА И ЕЕ СЛЕДСТВИЯ

При развитии космологии квантовую механику никогда не игнорировали. Свойства материи и излучения, спектральные линии, рассеяние света, статистики Бозе и Ферми — все эти вопросы учитывались при расчете давления, плотности энергии, спектральных коэффициентов переноса и т. д. Поэтому правая часть уравнений общей теории относительности Эйнштейна всегда включала квантовые эффекты.

Говоря об этих эффектах теперь, мы подчеркиваем влияние кривизны пространства-времени на частицы и поля в противоположность обычной физике пространства Минковского.

Наиболее интересным эффектом является рождение частиц гравитационным полем в вакууме. Реакции типа $e^{+}+e^{-}=g+g$, где g — гравитоны, были рассмотрены и рассчитаны в тридцатые и сороковые годы. Если берется много когерентных гравитонов, мы получаем классическую гравитационную волну. Рождение пар $e^{+}e^{-}$, очевидно, должно происходить в сталкивающихся пучках (классических) гравитационных волн, т. е. в вакууме с зависящей от времени метрикой.

В космологическом контексте возбуждение полей, т. е. рождение квантов поля (например, фотонов), в расширяющейся вселенной было впервые упомянуто Шредингером [58, 59] и позднее Утиямой и Де Виттом [65]. Подробное исследование было выполнено Паркером [50, 51]. Основной принцип состоит в том, что рождение частиц обусловлено неадиабатическим поведением соответствующего

поля в меняющейся метрике. Частицы рождаются с частотой порядка обратного характерного времени изменения метрики. Этот принцип приводит к порогу рождения массивных частиц: рождения нет при $t > \hbar/mc^2$. Согласно размерным аргументам, плотность энергии рожденных частиц порядка $\hbar/c^3t^4$. Важные качественные особенности рождения частиц заключаются в следующем.

1. Имеет место эффект поляризации вакуума — появление тензора энергии-натяжений без реальных частиц [1]). Одним из примеров является эффект Казимира [7] при нулевой температуре в статической ситуации. Здесь происходит чистая поляризация вакуума. В случае малого локализованного возмущения $g_{\mu\nu}=\eta_{\mu\nu}+h_{\mu\nu}$, где $h_{\mu\nu}=0$ при $t<t_1$ и $t>t_2$, $h_{\mu\nu}\ll 1$ и является гладкой функцией в интервале $t_1<t<t_2$. Поляризация вакуума представляет собой эффект первого порядка по $h_{\mu\nu}$ и исчезает вместе с возмущением метрики для $t>t_2$. Рождение реальных частиц является эффектом второго порядка $dm/dt \sim h^2$ (мы опускаем здесь тензорные индексы); эти частицы и их энергия и давление остаются после возмущения при $t>t_2$.

2. Второй важной особенностью теории является ее конформная инвариантность по крайней мере в пределе, когда массы покоя стремятся к нулю [9, 53]. Изотропное расширение конформно-эквивалентно статической ситуации.

Поэтому рождение частиц аномально мало (пропорционально $Gm^2/\hbar c = 10^{-38}$ в некоторой положительной степени) во фридмановских изотропных и однородных космологических моделях [2]).

С другой стороны [25], рождение частиц при анизотропном расширении приводит к расходящемуся результату, если оно включается при $t=0$. Если (искусственно) включить рождение при $t=t_{Pl}=$ $=10^{-43}$ с, то оно разрушает анизотропию в течение времени того же порядка [69, 78—80, 41]. Из этих соображений следует, что космологические решения, которые анизотропны в сингулярности, запрещены. Но мы имеем не только фридмановские модели. Квазиизотропные решения с изотропным расширением в пределе $t \to 0$ также возможны. Они имеют неоднородную метрику сопутствующего 3-пространства. Поэтому возмущения плотности, а также гравитационные волны развиваются из квазиизотропного решения. В то же время анизотропия возникает также, но достаточно поздно, чтобы

---

[1]) Важность этого эффекта для рождения частиц показана Зельдовичем и Питаевским [76].

[2]) Важным исключением являются гравитоны [26]. Общая теория относительности дает недвусмысленное классическое уравнение для гравитационных волн — с нулевой массой покоя, но еще не конформно-инвариантное. Чтобы предотвратить рождение гравитонов во фридмановской модели, требуется дополнительное условие обращения в нуль скаляра кривизны $R=0$, которое в свою очередь имеет место, если $\varepsilon={}^1/_3 p$ (где ε — плотность энергии и p — давление) в веществе фона. Возможно, что это условие выполняется в реальной вселенной вблизи сингулярности.

можно было бы избежать рождения частиц, если неоднородности метрики 3-пространства являются достаточно гладкими.

Самым важным и красивым результатом союза между квантовой теорией и гравитацией является теория испарения черных дыр Хокинга [27—29]. Ее физическое содержание рассматривается в более поздних главах этой книги. Упрощая, можно сказать, что коллапс в черную дыру — это событие, которое не оканчивается статической ситуацией (в отличие, например, от коллапса белого карлика в нейтронную звезду). Образование черной дыры приводит к экспоненциальному покраснению $\omega \sim e^{-ct/r_g}$ всего уходящего излучения. Поэтому для всех значений времени покраснение является неадиабатическим для уходящих волн с частотой $\omega \sim c/r_g$, поэтому коллапсирующая черная дыра испускает частицы с этой частотой. Этот расчет Хокинга, основанный на оценках порядка величин, был подтвержден доказательствами, связанными с детальным анализом характера излучения.

Испарение не важно для черных дыр звездной массы, оно важно только для первичных черных дыр (ПЧД). Понятие ПЧД впервые было сформулировано Зельдовичем и Новиковым [75].

Если начальная метрика и распределение плотности вблизи сингулярности достаточно неоднородны, то вполне возможно, что в каких-то частях первичной плазмы расширение сменяется сжатием, заканчивающимся ранним гравитационным коллапсом. Самоограничение квазиизотропными решениями не исключает возможности частичного гравитационного коллапса, т. е. образования ПЧД. Недавно Надежиным, Новиковым и Полнаревым [44] были получены численные примеры образования ПЧД из сферически-симметричных гладких квазиизотропных решений.

Рассматривая ПЧД в шестидесятые годы, Новиков и я не знали о квантовом испарении. Мы думали, что масса каждой отдельной ПЧД после ее образования не может уменьшаться; благодаря аккреции окружающего вещества она должна монотонно возрастать. Значение отношения конечной массы ПЧД к начальной массе было неопределенным; более поздняя аналитическая работа Карра и Хокинга [5] и еще более поздние расчеты Надежина и др. [44] показывают, что это отношение близко к единице. Но несмотря на отсутствие определенного значения этого отношения, с самого начала (1966—1967 гг.) было ясно, что начальная плотность ПЧД должна быть очень малой по сравнению с плотностью окружающей расширяющейся плазмы или, возможно, ПЧД не образуются вообще.

Следствием этого рассмотрения было утверждение, что начальная метрика сопутствующего трехмерного пространства является достаточно гладкой даже в самом малом масштабе. Этого утверждения нельзя было бы сделать, используя линейную теорию возмущений, так как в этой теории мелкомасштабные возмущения развиваются в колебания и полностью подавляются очень рано,

не оставляя наблюдаемых следов. Образование ПЧД находится вне области применимости линейной теории. Какие изменения в наших взглядах произошли за десять лет после открытия квантового испарения ПЧД?

Если масса ПЧД больше, чем планковская масса $(\hbar c/G)^{1/2} = 10^{-5}$ г, т. е. если коллапс ПЧД происходит позднее, чем планковское время 10^{-43} с, то образование и испарение разделены во времени, и испарение не влияет на образование. Связь между начальными возмущениями метрики и образованием ПЧД остается. Методы обнаружения и пределы обнаружения коренным образом меняются. ПЧД с начальной массой в интервале от $3\cdot10^{14}$ до 10^{16} г еще существуют до сих пор или исчезли недавно; они являются (или были) сильными излучателями жестких γ-фотонов, электрон-позитронных пар и мезонов. Такие очень активные объекты должны давать главный вклад в космические лучи, рентгеновский и гамма-фон [4, 49]. В упомянутом интервале масс чувствительность обнаружения активных (испаряющихся) ПЧД в 10^8 раз больше, чем чувствительность обнаружения пассивных ПЧД по их гравитации; предел наблюдаемости для испаряющихся ПЧД гораздо ниже.

ПЧД с временами испарения в интервале 100 с$<t<10^5$ лет (до рекомбинации) должны портить планковский спектр фонового излучения. ПЧД, которые испаряются раньше (в интервале 1 с$<$ $<t<$100 с), меняют характер нуклеосинтеза [66, 84]. При $t<1$ с ПЧД будут приводить к возрастанию энтропии; возможно, на самом деле часть энтропии обусловлена этим процессом, но нельзя получать больше, чем наблюдаемая энтропия на барион; вклад ПЧД должен быть меньше, чем 10^9 безразмерных единиц энтропии на барион.

Наши взгляды на космологическую важность ПЧД наинизшей массы, скажем от 10^{-5} до 1 г, испытали несомненные качественные изменения в свете достижений теории испарения ПЧД. В шестидесятые годы они казались запрещенными даже более сильно, чем тяжелые ПЧД.

Теперь в связи с теорией испарения можно допустить, что их обильное начальное образование сопровождало очень раннее испарение [1]). Сильные отклонения от однородности допускаются в самых малых масштабах — от планковского 10^{-33} см при $t=10^{-43}$ с до 10^{-28} см при $t=10^{-38}$ с (это соответствует масштабу от $\sim10^{-2}$ до 10^3 см после расширения).

Образование и испарение ПЧД ведет к эффективному нарушению закона сохранения барионов. Имеется тривиальная возможность, состоящая в том, что энергия исчезающих барионов может использоваться для снабжения теплом и энтропией остающихся барионов [6].

---

[1]) Подробный обзор допустимого количества ПЧД дан в докладе Новикова, Старобинского и Зельдовича [45] на конференции по ОТО (GR VIII) в Ватерлоо, Канада, в августе 1977 г.

Однако в этом случае ПЧД не являются специфическим средством: возможно, по крайней мере в принципе, построить начальную сингулярность с возмущениями метрики, достаточно сильными, чтобы дать необходимую энтропию без образования ПЧД и(или) несохранения барионов. В частном случае жесткого уравнения состояния $p=\varepsilon$ это было показано автором [70, 71].

Другая, нетривиальная возможность заключается в том, чтобы получить наблюдаемую малую асимметрию барионного заряда ($(B-\bar{B})/\gamma \sim 10^{-8}$) из полностью симметричной начальной ситуации. Лабораторными экспериментами твердо установлено существование нарушения CP-инвариантности, т. е. абсолютное различие некоторых свойств (относительные ширины распадов, но не массы) частиц и античастиц. Но для получения избытка барионов из симметричного состояния также требуется несохранение барионного заряда. В более ранних схемах [36, 56], где оно вводилось на кварково-частичном уровне, трудно было согласовать эту гипотезу с устойчивостью обычного вещества (см. [52] — расхождение!).

Хокинг указал, что ответ может дать образование и испарение ПЧД. Зельдович [72] попытался дать конкретный вариант этой идеи. Само испарение зарядово-симметрично, но если нестабильные частицы и античастицы, возникающие при испарении, имеют различные свойства распада, то неполная аккреция продуктов распада может приводить к асимметричной вселенной. На самом деле эта гипотеза очень спекулятивна. Все процессы предполагаются происходящими на ранней стадии адронной эры.

Самая последняя проблема, возможно связанная с квантованием метрики, касается «больших чисел» $\hbar c/Gm_p^2=10^{37}$ или $\hbar c/Gm_e^2=10^{43}$, которые получаются при сопоставлении физики частиц с гравитацией.

В замечательной работе Дирака [15] было сделано отрицательное утверждение, что такие числа невозможно получить ни в какой разумной локальной теории. Позитивное утверждение Дирака состояло в том, что большие числа связаны с космологическими большими числами, например с квадратным корнем $\sqrt{N}$ из числа частиц внутри горизонта $N=(c/H)^3 n \sim 10^{80}$ или с отношением возраста вселенной $t \sim 10^{18}$ с к характерному адронному времени $\tau=\hbar/m_p c^2 = 10^{-24}$ с: $t/\tau=10^{42}$. Эти грубые совпадения, если рассматривать их всерьез, говорят о нелокальной, маховской физике. Они также подразумевают, что физические «константы», или по крайней мере одна из них, зависят от времени. Обычно под подозрением оказывается G.

Отрицательное утверждение, конечно, справедливо для обычных теорий типа теории возмущения. В этих теориях безразмерные константы являются степенями 2, 3, π, $\hbar c/e^2=137$ и т. д., и утверждение Дирака является убедительным. Разумный человек не будет строить теорию с $(\hbar c/e^2)^{18}$.

Однако в последние несколько лет появились примеры теорий нового типа, главным образом благодаря Полякову [55] и его коллегам [2]. Изменения полевой топологии связаны с промежуточными состояниями, не подчиняющимися классическим уравнениям. Один эффект связан с проникновением через барьер (туннелированием). В таких случаях появляются экспоненциально малые безразмерные числа, например $e^{-16\pi^2/g^2} \sim 10^{-430}$.

В гравитации не существует безразмерных зарядов. Тем не менее можно получать числа типа $e^{8\pi^2} = 6 \cdot 10^{30}$. Например, можно строить предположения, основанные на существовании «кротовых нор» (топологических ручек [1]), делающих пространство неориентируемым: частица с правосторонней спиральностью падает в «кротовую нору», частица с левосторонней спиральностью появляется. Топологическая ручка играет роль массы покоя в дираковской теории частиц со спином $^1/_2$, связывая две различные спиральности. В пространстве, усеянном виртуальными топологическими ручками такого типа, каждая частица со спином $^1/_2$ будет приобретать массу покоя. Первая оценка дает значение $m = (\hbar c/G)^{1/2} = 10^{-5}$ г (просто из соображений размерности), но если бы топологические ручки «туннелировали», то можно было бы предположить наличие экспоненциально малого безразмерного коэффициента, приводящего массы частиц в приемлемый диапазон значений. Я понимаю, что частный пример «кротовой норы» является достаточно сомнительным и не свободным от трудностей (нейтрино с нулевой массой покоя и нарушением четности [75]). Но общее утверждение, что новые типы теорий могли бы дать экспоненциально малые (или большие) числа, по-видимому, верно. Космологическое маховское объяснение больших чисел уже не является необходимым [74].

Развитие общей теории относительности носит осциллирующий характер. Эйнштейн начал с кривизны пространства-времени. Много усилий было затрачено на переформулирование теории в виде более привычной нелинейной полевой теории для тензора второго ранга в плоском пространстве. Космология с ее характерным интересом к замкнутым мирам всегда была бастионом геометрического подхода. В последнее десятилетие, когда интерес сместился в сторону черных дыр и «кротовых нор», еще отчетливее выступили красота и плодотворность геометрических идей; но, когда их смешивают с квантовой теорией, предсказания становятся рискованными. Qui vivra, verra! [2]

---

[1]) Идею о топологических ручках как носителях электрического заряда пропагандировал Уилер [68].

[2]) Поживем — увидим! (*франц.*).— *Прим. перев.*

ЛИТЕРАТУРА

1. *Alpher R., Herman R.*, Ann. Rev. Nucl. Sci., **2**, 1 (1953).
2. *Belavin A. A. et al.*, Phys. Lett., **B59**, 85 (1975).
3. *Богданова Л. М., Шапиро И. С.*, Письма в ЖЭТФ, **20**, 217 (1974).
4. *Carr B. J.*, Astrophys. J., **206**, 8 (1976).
5. *Carr B. J., Hawking S. W.*, Mon. Not. Roy. Astron. Soc., **168**, 399 (1974).
6. *Carr B. B., Rees M.*, preprint (1977).
7. *Casimir H. B. C.*, Proc. Nederl. Acad. Wetensch., **60**, 793 (1948).
8. *Chapline G. F.*, Nature, **253**, 1 (1975).
9. *Chernikov N. A., Tagirov E. A.*, Ann. Inst. H. Poincaré, **9**, 109 (1968).
10. *Chiu H. Y.*, Phys. Rev. Lett., **17**, 712 (1966).
11. *Coleman S.*, Phys. Rev., **D15**, 2929 (1977).
12. *Cowsik R., McClelland J.*, Phys. Rev. Lett., **29**, 669 (1972).
13. *Dicke R. H., et al.*, Astrophys. J., **142**, 414 (1965).
14. *Dicus D. A., Kolb E. M., Teplitz V. L.*, Phys. Rev. Lett., **39**, 168 (1977).
15. *Dirac P. A. M.*, Nature, **139**, 323 (1937). [Имеется перевод: Альберт Эйнштейн и теория гравитации.— М.: Мир, 1979, с. 538.]
16. *Дорошкевич А. Г., Новиков И. Д.*, ДАН СССР, **154**, 809 (1964).
17. *Einstein A.*, Sitzungsber. Preuss. Akad. Wiss., **1**, 142 (1917). [Имеется перевод: СНТ, т. 1, 601.— М.: Наука, 1965.]
18. *Frampton P. H.*, Phys. Rev., **D15**, 2922 (1977).
19. *Friedman A.*, Z. Physik, **10**, 377 (1922).
20. *Friedman A.*, Z. Physik, **21**, 326 (1924).
21. *Gamow G.*, Phys. Rev., **70**, 572 (1946).
22. *Gamow G.*, Phys. Rev., **86**, 251 (1952).
23. *Gamow G.*, In Vistas in Astronomy, ed. X. Beer, vol. 2, p. 1726. Pergamon Press: Oxford, New York, 1956.
24. *Герштейн С. С., Зельдович Я. Б.*, Письма в ЖЭТФ, **4**, 174 (1966).
25. *Гриб А. А., Мамаев С. Г.*, Ядерная физика, **10**, 1276 (1969).
26. *Грищук Л. П.*, ЖЭТФ, **67**, 825 (1974).
27. *Hawking S. W.*, Nature, **248**, 30 (1974).
28. *Hawking S. W.*, Commun. Math. Phys., **43**, 199 (1975).
29. *Hawking S. W.*, Quantum gravity and path integrals. (Talk given at the 8th International Conference on General Relativity and Gravitation, Waterloo, Ontario, August 9th (1976).
30. *Hayashi C.*, Progr. Theor. Phys., **5**, 224 (1950).
31. *Киржниц Д. А.*, Письма в ЖЭТФ, **15**, 745 (1972).
32. *Kirzhnitz D. A., Linde A. D.*, Phys. Lett., **B42**, 471 (1972).
33. *Киржниц Д. А., Линде А. Д.*, ЖЭТФ, **67**, 1263 (1974).
34. *Kirzhnitz D. A., Linde A. D.*, Ann. Phys. (N. Y.), **101**, 195 (1976).
35. *Кобзарев И. Ю., Окунь А. Б., Волошин М. Б.*, Ядерная физика, **20**, 1229 (1974).
36. *Кузьмин В. А.*, Письма в ЖЭТФ, **12**, 335 (1970).
37. *Lee B. W., Weinberg S.*, In Fermilab-Pub-77/41—THY, (1977).
38. *Lee B. W., Weinberg S.*, Phys. Rev. Lett., **39**, 165 (1977).
39. *Lee T. D.*, Phys. Rev., **D8**, 1226 (1973).
40. *Линде А. Д.*, Письма в ЖЭТФ, **19**, 320 (1974).
41. *Лукаш В. Н., Старобинский А. А.*, ЖЭТФ, **66**, 1515 (1974).
42. *McKellar A.*, Publ. Dominion Astrophys. Observ., **7**, No. 15 (1941).
43. *Milne E.*, Q. J. Math., **5**, 64 (1934).
44. *Надежин Д. К., Новиков И. Д., Полнарёв А. Г.*, АЖ, **55**, 2 (1978).
45. *Novikov I. D., Starobinsky A. A., Zel'dovich Ya. B.*, Primordial Black Holes. 8th International Conference on General Relativity and Gravitation, Waterloo, Ontario, 1977.
46. *Omnes R.*, Astron. Astrophys., **10**, 228 (1971).
47. *Omnes R.*, Astron. Astrophys., **11**, 450 (1971).
48. *Omnes R.*, Astron. Astrophys., **15**, 273 (1971).
49. *Page D. K., Hawking, S. W.*, Astrophys. J., **206**, 1 (1976).

50. *Parker L.*, Phys. Rev. Lett., **21**, 562 (1968).
51. *Parker L.*, Phys. Rev., **183**, 1057 (1969).
52. *Pati J. S., Salam A.*, Phys. Rev. Lett., **31**, 661 (1973).
53. *Penrose R.*, In Relativity, Groups and Topology, eds. C. M. DeWitt and B. S. DeWitt, p. 565. Gordon and Breach, New York, 1964.
54. *Penzias A. A., Wilson R. W.*, Astrophys. J., **142**, 419 (1965).
55. *Polyakov A. M.*, Phys. Lett., **B59**, 78 (1975).
56. *Сахаров А. Д.*, Письма в ЖЭТФ, **5**, 32 (1967).
57. *Sato K., Sato H.*, Prog. Theor. Phys., **54**, 912 (1975).
58. *Schrödinger E.*, Physica, **6**, 899 (1939).
59. *Schrödinger E.*, Proc. R. Irish Acad., **46**, 25 (1970).
60. *Шварцман В. Ф.*, Письма в ЖЭТФ, **9**, 315 (1969).
61. *Steigman G.*, In General Relativity and Cosmology, ed. R. K. Sachs. Academic Press, London, New York, 1971.
62. *Steigman G.*, In Proc. IAU Symp., No. 63, Poland, 1973.
63. *Steigman G., Schramm D. N., Gunn J. E.*, Phys. Lett., **B66**, 202 (1977).
64. *Szalay A. S., Marx G.*, Astron. Astrophys., **49**, 437 (1976).
65. *Utiyama R., DeWitt B. S.*, J. Math. Phys., **3**, 608 (1962).
66. *Вайнер Б. В., Насcельский П. Д.*, Письма в АЖ, **3**, 147 (1977).
67. *Высотский М. И., Долгов А. Д., Зельдович Я. Б.*, Письма в ЖЭТФ, **26**, 200 (1977).
68. *Wheeler J. A.*, Geometrodynamics. Academic Press: London, New York, 1962.
69. *Зельдович Я. Б.*, Письма в ЖЭТФ, **12**, 443 (1970).
70. *Zel'dovich Ya. B.*, Mon. Not. R. Astron. Soc., **160**, 1L (1972).
71. *Зельдович Я. Б.*, ЖЭТФ, **64**, 58 (1973).
72. *Зельдович Я. Б.*, Письма в ЖЭТФ, **24**, 29 (1976).
73. *Zel'dovich Ya. B.*, Annu. Rev. Fluid Mech., **9**, 215 (1977).
74. *Зельдович Я. Б.*, ИФН, **123**, 487 (1977).
75. *Зельдович Я. Б., Новиков И. Д.*, АЖ, **44**, 663 (1967).
76. *Zel'dovich Ya. B., Pitaevsky L. P.*, Commun. Math. Phys., **23**, 185 (1971).
77. *Зельдович Я. Б., Смородинский Я. А.*, ЖЭТФ, **41**, 907 (1961).
78. *Зельдович Я. Б., Старобинский А. А.*, ЖЭТФ, **61**, 2161 (1972).
79. *Зельдович Я. Б., Старобинский А. А.* в сб. «Проблемы ядерной физики и физики элементарных частиц», с. 141, «Наука», М., 1975.
80. *Зельдович Я. Б., Старобинский А. А.*, Письма в ЖЭТФ, **24**, 616 (1976).
81. *Зельдович Я. Б., Окунь Л. Б., Пикельнер С. Б.*, ИФН, **87**, 113 (1965).
82. *Зельдович Я. Б., Илларионов А. Ф., Сюняев Р. А.*, ЖЭТФ, **62**, 1217 (1972).
83. *Зельдович Я. Б., Кобзарев И. Ю., Окунь Л. Б.*, Препринт ИПМ № 15, 1974; ЖЭТФ, **67**, 3 (1974).
84. *Зельдович Я. Б. и др.*, Письма в АЖ, **3**, 208 (1977).

V. СИНГУЛЯРНОСТИ И АСИММЕТРИЯ ПО ВРЕМЕНИ

Р. Пенроуз [1])

1. ВВЕДЕНИЕ

Общая теория относительности, представляющая собой в высшей степени стройное описание геометрии мира, приводит к пространственно-временно́й картине, в которой, по-видимому, неизбежны сингулярности, и этот факт остается для многих источником беспокойства. Сам Эйнштейн боролся против таких кажущихся изъянов своей теории и с немалой изобретательностью наметил ряд возможных выходов из положения (например, «мост» Эйнштейна — Розена [1], попытка построения модели устойчивых, не подверженных коллапсу звездных скоплений [2], идея о возможности несингулярного «схлопывания» Вселенной за счет неоднородностей [3], вплоть до его попыток видоизменить общую теорию относительности таким образом, чтобы получилась свободная от сингулярностей единая теория поля [4, 5]). Тем не менее исследования более позднего времени все настойчивее склоняют нас к тому, чтобы примириться с существованием сингулярностей и принять их как исконное свойство геометрии Вселенной.

Это не означает, что неотъемлемой частью нашего описания физической геометрии должно стать теперь некоторое математически строгое понятие «сингулярности», хотя теоретики много и искусно поработали в этом направлении за последние годы. По-видимому, дело скорее в том, что ограничена область применимости самого понятия геометрии пространства-времени и соответственно физических законов в их нынешнем понимании. Основываясь на теоремах о сингулярностях [6—8], можно даже сказать, что эти законы *сами себя* ограничивают. Однако мне не кажется, что это повод для пессимизма. Так или иначе, возникла необходимость в каких-то новых законах, и в то же время можно думать, что существование и вероятная структура пространственно-временны́х сингулярностей содержат ключ к решению одной из самых древних загадок физики — откуда берется *стрела времени*?

Точка зрения, которую я хочу изложить, возникла вовсе не от какого-то радикального взгляда на вещи. Как бы ни казалась суть моих конечных выводов отличной от того, что по этому предмету обычно говорят, я все же придерживаюсь (или, как мне казалось, придерживаюсь) общепринятого подхода к большинству вопросов.

---

[1]) *R. Penrose*, University of Oxford, Mathematical Institute, Oxford, UK.

Мои доводы основаны не на детальных вычислениях, а на некоторых, как мне кажется, «очевидных» фактах, сама очевидность которых способствует тому, что их обычно не принимают во внимание.

Именно Эйнштейн, как никто другой, владел искусством получения глубоких физических результатов на основе «очевидных» фактов. Надеюсь, что меня можно простить за попытку соперничать с Эйнштейном в его столетний юбилей, ибо я уверен, что он сам не очень высоко оценил бы мнимую значительность столь условной годовщины.

2. СУТЬ ПРОБЛЕМЫ

Главный вопрос всем хорошо знаком [9—11]. Локальные физические законы, известные и понятные нам, все симметричны по времени. И все-таки на макроскопическом уровне наблюдается явная асимметрия по времени. Действительно, можно указать ряд на первый взгляд различных макроскопических стрел времени. К ним можно добавить единственную асимметрию по времени, наблюдающуюся в физике фундаментальных частиц, — она проявляется в распаде K^0-мезона. Кроме того, сюда же относится вопрос об интерпретации квантовой механики; мне кажется, что его не следует неосмотрительно упускать из виду. Установилось мнение, что в аппарате квантовой механики на самом деле нет никакой стрелы времени [9, 12—16], хотя на первый взгляд кажется, что она там есть. Я не оспариваю это мнение, но тем не менее считаю, что данный вопрос не должен быть предан забвению по причинам, которые я изложу позднее.

Итак, я позволю себе перечислить семь, по-видимому, независимых стрел (иные из них являются лишь возможными), которые обсуждаются в литературе [9]: распад K^0-мезона (разд. 2.1), квантовомеханические наблюдения (разд. 2.2), общий рост энтропии (разд. 2.3), запаздывание излучения (разд. 2.4), психологическое время (разд. 2.5), расширение Вселенной (разд. 2.6) и соотношение черных и белых дыр (разд. 2.7). Рассмотрим их по порядку.

2.1. РАСПАД K^0-МЕЗОНА

Может ли асимметрия, которая появляется в распаде K^0-мезона, иметь хотя бы отдаленную связь с другими стрелами? Ведь этот эффект в сущности ничтожен. T-нарушающая компонента в этом распаде составляет, вероятно, всего лишь 10^{-9} от T-сохраняющей компоненты [17—21], и, во всяком случае, существование T-нарушения обнаруживается не непосредственным измерением, а выводится из наличия незначительного CP-нарушения ($\sim 10^{-9}$) в предположении, что CPT-нарушение, если оно вообще существует, должно быть в этом взаимодействии еще меньше ($\ll 10^{-9}$). По-видимому, здесь замешано сверхслабое взаимодействие (или очень слабая

компонента слабого взаимодействия), хотя оно не играет, насколько нам известно, никакой заметной роли в каких-либо важных процессах, управляющих поведением материи. Гравитация, конечно, еще слабее, и притом она полностью управляет движением материи в больших масштабах. Однако дифференциальные уравнения общей теории относительности полностью симметричны по времени, подобно уравнениям Максвелла, а также, по-видимому, подобно законам сильных и обычных слабых взаимодействий.

Так или иначе, реальное, но ничтожное проявление почти полностью скрытой асимметрии по времени, по-видимому, имеется в K^0-распаде. Трудно поверить, что Природа, образно говоря, не «пытается нам что-то сообщить» в результатах этого тонкого и красивого эксперимента, который был неоднократно подтвержден [20]. Одно из ранних предположений сводилось к тому, что T-нарушающий эффект возникает в связи с дальнодействующим космологическим взаимодействием, вследствие которого асимметрия между веществом и антивеществом приводит к требуемой асимметрии по времени [50]. Однако при последующем анализе [19, 20] эта точка зрения была сочтена неприемлемой. Вероятно, эта асимметрия действительно содержится в *локальных* динамических законах. К этому вопросу я еще собираюсь вернуться, поскольку уверен в его важности.

2.2. КВАНТОВОМЕХАНИЧЕСКИЕ НАБЛЮДЕНИЯ

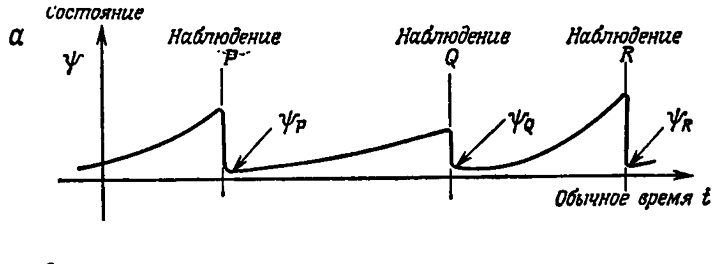
В стандартной квантовой механике динамическая эволюция состояния происходит согласно уравнению Шредингера. При обращении времени это уравнение преобразуется в себя при условии, что i заменяется на $-i$. Однако уравнение Шредингера должно быть дополнено процедурой («редукция волновой функции»), в соответствии с которой как только над системой производится «наблюдение», текущий волновой вектор отбрасывается и заменяется собственным состоянием ψ_Q эрмитова оператора Q, представляющего наблюдаемую, которая измерялась. На первый взгляд эта процедура асимметрична по времени, поскольку состояние системы является собственным состоянием оператора Q сразу после проведения наблюдения, но оно (обычно) не является таковым непосредственно перед наблюдением (рис. 1, *а*). Однако эта асимметрия описания легко устраняется [14]: при описании, обращенном по времени, к этому же самому состоянию ψ_Q можно просто относиться как к состоянию непосредственно перед измерением, а уравнение Шредингера использовать для задания движения состояния в прошлое до момента предыдущего измерения (отвечающего оператору P). Затем этот развивавшийся в прошлое вектор состояния отбрасывается (в соответствии с обращенной по времени версией «редукции волновой функции») и подставляется собственное состояние ψ_P оператора P

(соответствующее реальному результату наблюдения величины P) (рис. 1, *б*).

Относительная вероятность наблюдения определенного значения Q при данном значении P или наблюдения определенного значения P при данном значении Q одна и та же при каждом из двух возможных описаний и равна

$$|\langle\psi_Q | U\psi_P\rangle|^2 = |\langle\psi_P | U^{-1}\psi_Q\rangle|^2,$$

где U — унитарный оператор, определяющий эволюцию состояния (согласно уравнению Шредингера) с момента измерения P



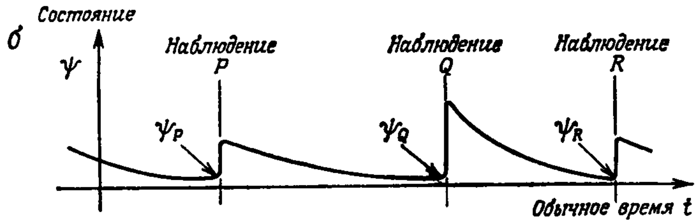


Рис. 1. *а* — общепринятая (шредингеровская) картина развития волновой функции; *б* — существенно эквивалентное «обращенное по времени» развитие волновой функции.

до момента измерения Q. Поскольку эти вероятности в любом случае — единственные наблюдаемые проявления волновой функции ψ, мы видим, что эти две возможности описания эквивалентны и что аппарат квантовой механики симметричен по времени.

Часто встречается утверждение [16, 22], что значения волновой функции в некоторый момент времени сами по себе не дают полного описания физической реальности. Это весьма наглядно иллюстрируется рис. 1, если только симметрия по времени действительно свойственна микрофизике. Но безоговорочное принятие точки зрения, будто волновая функция полностью описывает физическую реальность, в любом случае сталкивается с общеизвестной трудностью, которая возникает уже в специальной теории относительности. Дело в том, что «наблюдение» редуцирует волновую функцию в одно из своих собственных состояний, по-видимому, одновременно с самим наблюдением, причем слово «одновременно», надо полагать, относится к системе покоя наблюдателя. При проведении двух про-

странственноподобно разделенных наблюдений это может привести к концептуальным проблемам, связанным с вопросом о *последовательности*, в которой происходят две редукции волновой функции. Эта трудность особенно ярко проявляется в знаменитом мысленном эксперименте Эйнштейна — Подольского — Розена [22].

Однако имеются *иные* ситуации, в которых, по всей видимости, трудно придерживаться точки зрения, будто волновая функция (или вектор состояния) *не дает* подлинного описания физической реальности в соответствии с рис. 1, *а*. Рассмотрим изолированную систему, над которой только что произведено наблюдение P, дающее *общепринятое* описание некоторого состояния ψ_P (собственное состояние оператора P); это состояние развивается вперед во времени согласно уравнению Шредингера, что приводит через некоторое время к состоянию $U\psi_P$. Теперь (в рамках общепринятой квантовой механики и в предположении, что никакие дополнительные принципы не привлекаются) оператор UPU^{-1} представляет собой столь же «хорошую» наблюдаемую, что и P. Более того, в предположении, что собственное значение λ, соответствующее ψ_P, не вырождено, состояние $U\psi_P$ обладает тем свойством, что оно (возможно, с точностью до множителя, отличного от нуля) является *единственным* состоянием, для которого вероятность получения значения λ при наблюдении величины UPU^{-1} равна единице. Изолированная система не может (при обычном взгляде на вещи) «знать», что над ней, возможно, собираются произвести наблюдение UPU^{-1}, но она должна быть готова к такой неожиданности! Таким образом, информация о состоянии $U\psi_P$ (с точностью до фазы) должна быть запасена в самой системе, т. е. волновая функция *в самом деле* описывает физическую реальность.

Конечно, UPU^{-1} может соответствовать чересчур диковинному и совершенно невыполнимому эксперименту, как, например, если бы в случае шредингеровского «парадокса кошки» [24] нам нужно было убедиться, что конечное состояние $U\psi_P$ есть «комплексная линейная комбинация дохлой кошки и живой кошки». Сама диковинность такого рода экспериментов наводит на мысль, что $U\psi_P$ (и UPU^{-1}) может в конце концов и не иметь «реального» отношения к действительности. Однако это моя личная точка зрения. В законах квантовой механики при применении их к подобным ситуациям чего-то недостает или проступает какое-то несоответствие. Есть также какая-то несуразность в самой идее редукции волновой функции как способа описания физической реальности (это же относится к любому другому альтернативному, но существенно эквивалентному способу описания этого процесса, как, например, описанию хитроумным наблюдателем своего пути, который он прокладывает сквозь многолистную Вселенную Эверетта [21—29]). Но для чего же еще физическая теория, как не для описания физической реальности? Мне кажется, что в этом вопросе я должен равняться на тех основоположников квантовой механики (и не в последнюю очередь

на самого Эйнштейна), которые были убеждены, что решение проблемы «наблюдений» не может быть найдено в рамках самой квантовой механики [1]). Итак, по-видимому, требуется какая-то новая (предположительно нелинейная) теория, из которой квантовая механика и классическая механика вытекают как различные предельные случаи.

Рассматриваемый вопрос важен для нас по двум причинам. Во-первых, осуществление наблюдения, по всей видимости, связано с каким-то *необратимым* процессом и предполагает существенный рост энтропии. Вовсе не очевидно, что недостающим (или непонятным) звеном в квантовой механике не является некий фундаментальный, но *асимметричный по времени* закон. Поэтому демонстрация симметрии по времени формализма квантовой механики на самом деле не закрывает вопроса об (а)симметрии по времени в квантово-механических измерениях.

Вторая причина, по которой данный вопрос важен для нас, касается роли квантовой гравитации. Нельзя упускать из виду возможность, что присутствие (измеримого) *гравитационного поля* — это как раз то, что выводит описание физической системы за рамки чисто квантовой физики [32, 33]. А если необходим *новый закон*, особенно для понимания ситуаций, в которых и квантовая, и классическая физика доходят до своих пределов, тогда плохи ваши дела, квантователи гравитации!

Я вернусь к этим вопросам в разд. 4. А пока в своем отношении к квантовой механике я намерен придерживаться вполне стандартного взгляда: в квантовой механике нет явной стрелы времени, и решение проблемы макроскопической асимметрии по времени следует искать где-то в другом месте.

2.3. ОБЩИЙ РОСТ ЭНТРОПИИ

Для обсуждения асимметрии по времени ключевым является, конечно, статистическое понимание энтропии. И если все (важнейшие) локальные законы симметричны по времени, то источник статистических асимметрий следует искать в граничных условиях. При этом форма локальных законов предполагается такой, чтобы они, подобно ньютоновской теории, обычной теории Максвелла — Лоренца, гамильтоновой теории, уравнению Шредингера и т. д., определяли эволюцию системы, коль скоро заданы граничные условия, и, кроме того, чтобы эти граничные условия было достаточно задать *либо* в прошлом, *либо* в будущем. (Граничные значения обычно задаются на пространственноподобной гиперповерхности.) Тогда появление статистической стрелы времени могло быть обусловлено тем фактом, что *начальные* граничные условия непременно имеют намного *меньшую* энтропию, чем *конечные* граничные условия.

[1]) Достаточно упомянуть Бора [30], Шредингера [24], Дирака [21], Вигнера [25].

Прежде чем двигаться дальше, следует обсудить ряд вопросов. Во-первых, в определении поведения системы путем задания граничных условий в прошлом *или* будущем вообще есть что-то не вполне разумное. Эта «неразумность» особенно отчетливо видна в случае граничных условий в будущем. Допустим, я роняю мои часы на каменный пол и они разлетаются вдребезги, а затем жду 10 минут. Граничными условиями в будущем является куча колесиков и пружин, но с таким детальным и непостижимо точным подбором скоростей, что, будучи обращены назад (т. е. при обратном отсчете времени), они внезапно воссоздают мои часы после 10-минутного периода видимой неподвижности. Хотя модели физики, которыми мы пользуемся (например, ньютоновская теория), и допускают в принципе задание положений и скоростей с такой точностью, ниоткуда не следует (и мы этого не требуем), чтобы наши модели физических законов с такой точностью соответствовали действительности. Эта же проблема, как подчеркивал Борн [34], существует и для прошлых граничных условий. Фейнман [35] отмечал, что в ньютоновской теории, если все положения и скорости сложной системы известны с определенной точностью, то вся эта точность теряется за время меньшее, чем нужно для того, чтобы выразить эту точность словами! Борн (и Фейнман) прибегли к этому аргументу, чтобы продемонстрировать, что в классической механике в некотором смысле детерминизма не больше, чем в квантовой. Конечно, в квантовой механике есть дополнительная проблема: что происходит, когда совершается «наблюдение»? «Наблюдения» же представляются необходимыми с нормальной точки зрения, чтобы удержать волновые функции от расплывания по всему пространству.

Я упомянул об этих вещах, в основном чтобы отметить трудности. Но впредь мы будем их игнорировать и следовать общепринятому взгляду, что граничные условия действуют! Есть, однако, в какой-то мере связанный с этим вопрос, который требует дальнейших замечаний. Рассмотрим снова мои разбитые часы как граничные условия в будущем. Можно видеть, что хотя это состояние обладает более высокой энтропией, чем состояние, существовавшее до того, как часы разбились,— и потому при нормальном взгляде на вещи оно является не столь «необычным» состоянием, чем более раннее состояние в прошлом,— тем не менее более позднее состояние, если исследовать его в мельчайших подробностях, все же очень странное ввиду чрезвычайно точных корреляций между движениями отдельных частей системы. И снова я буду придерживаться здесь общепринятого «макроскопического» взгляда. Эта странность не того сорта, о которой можно сказать, что «энтропия мала». И «причина» того, что у меня раньше были часы, не в том, что существуют эти точные корреляции в *будущих* граничных условиях, а в том, что нечто в *прошлом* (скажем, часовой завод) имело более низкую энтропию, чем оно могло бы иметь в любом другом случае.

Подобным же образом и «причина» точных корреляций в движениях частей разбитых часов может быть прослежена обратно к заводу, но ни в какое другое место.

Мне не кажется, что этим и решается вопрос асимметрии по времени. Вполне возможно, что в какой-то соответствующим образом устроенной вселенной одни процессы могли бы происходить подобно истории с моими часами, а другие (по смыслу того самого времени, которое показывали мои часы, когда они еще шли!) были бы подвержены таким чудесным согласованным превращениям, что обеспечить «причиной» *их* поведение, наверное, можно было бы, только привлекая специфические (низкоэнтропийные) граничные условия в *будущем*. Но не похоже, чтобы так была устроена наша Вселенная.

Я до сих пор затушевывал относящийся сюда важный вопрос о крупнозернистой структуре [36]. Что понимать под энтропией? Является ли она определенным физическим атрибутом данной системы, который, подобно энергии-импульсу, очевидно, не зависит от того, как мы смотрим на систему? На практике энтропию можно свободно трактовать подобным образом (например, в физической химии). Но для наиболее общих определений энтропии, от которых можно ожидать, что они применимы к сложным системам вроде моих часов, нам необходим, по-видимому, довольно произвольный (т. е. необъективный) способ объединения физических состояний в более крупные классы (крупнозернистую структуру), в которых элементы каждого класса считаются неразличимыми между собой. При этом понятие энтропии относится не к отдельным состояниям, а к классам состояний. Тогда энтропия (по Больцману) класса, содержащего N различных (квантовых) состояний, равна

$$S = k \ln N,$$

где k — постоянная Больцмана. На самом деле, в нашем распоряжении имеется несколько определений энтропии [36—38]. Но в целом данный вопрос, несомненно, изобилует трудностями [1]). (Явление «спинового эха» служит одним из впечатляющих примеров, отражающих эти трудности [39]. Я даже не уверен, что «рост энтропии» вообще является подходящим понятием для описания того, как разбились мои часы. Вероятно, принятие ванны увеличивает энтропию значительно больше, чем в случае моих часов, в котором сопоставимый рост энтропии должен быть совсем незначительным.) К вопросу об объективном характере энтропии мы вернемся в разд. 4. А пока я на время еще раз поспешно уйду в убежище общепринятого мнения: энтропия — это понятие, которым можно жонглировать самым бесцеремонным образом!

---

[1]) К тому же я оставляю в стороне такие важные вопросы, как «H-теорема» [36 — 38] и «ветвящиеся системы» [10, 11]. Они *не связывают* причину роста полной энтропии с асимметрией по времени.

2.4. ЗАПАЗДЫВАЮЩЕЕ ИЗЛУЧЕНИЕ

Вопрос о граничных условиях имеет близкое отношение к следующей из наших стрел — к стреле запаздывающего излучения. Мы можем рассматривать это явление в двух совершенно различных аспектах, а именно: снова возвращаясь к вопросу об энтропии и, с другой стороны, в связи с вопросом об излучении без источников или без стоков. Запаздывание, конечно, не является свойством одного только электромагнитного излучения, хотя и рассматривается, как правило, именно в таком контексте. Представим себе, что в пруд брошен камень. Мы ожидаем увидеть, что круги, расходящиеся от точки падения, постепенно потеряют энергию в результате диссипации, особенно при ударе о берег. Мы вовсе не ожидаем увидеть, что еще до достижения камнем воды у берегов возникнут волны, столь тщательно организованные, что они сойдутся именно в точке входа камня в воду точно в тот момент, когда он коснется воды. Еще менее того мы рассчитываем, что как только эти круги сойдутся от берегов в некоторой точке посреди пруда, из этой точки вверх взлетит камень! Но такое экстраординарное поведение находится в полном согласии с локальными физическими законами. Однако для того, чтобы оно осуществилось, потребовались бы такие точные корреляции в движениях частиц, которые можно было бы объяснить только какими-то низкоэнтропийными граничными условиями в будущем.

Считаю необходимым подчеркнуть еще раз одно из положений предыдущего раздела, поскольку мне думается, что это ключевой момент: корреляции в деталях поведения частиц в будущем присутствуют *потому*, что в прошлом энтропия была мала. Аналогично такие же (но обращенные во времени) корреляции отсутствуют в прошлом *потому*, что энтропия велика в будущем. Последнее утверждение звучит необычно, но я стараюсь не связывать себя с упорядочением по времени [1]). Моя точка зрения состоит в том, что эти корреляции нельзя рассматривать как «причину» чего-либо вообще; но низкая энтропия (которая сама должна быть объяснена какой-то другой причиной) *может* служить «причиной» корреляций. (Таким образом, мы обходим проблему беспредельной точности физических законов.) Я еще раз подчеркну, что система, «специальная» в смысле наличия у нее такого рода сложных корреляций между частицами, *не* является «специальной» в том смысле, что *в этот момент времени* ее энтропия мала. Это самый существенный пункт в вопросе о крупнозернистой структуре.

Таким образом, мы видим, что нормальное запаздывающее поведение ряби соответствует низкой энтропии в прошлом и корреля-

[1]) Это ведет к логическому обращению точки зрения, выраженной О. Пенроузом и Персивалем [40] в их «законе условной независимости», согласно которой отсутствие начальных корреляций нужно *постулировать*. Картина мира, представленная в разд. 3.3, служит определенным подтверждением этого закона.

циям в будущем; в то же время описанные мной две ситуации, в которых происходит как бы *опережающее* движение ряби, связаны с очень точными корреляциями такого сорта, который ведет к снижению энтропии. Более того, альтернативная гипотетическая ситуация с *запаздыванием*, в которой камень внезапно вылетает из пруда, сопровождаемый рябью, расходящейся к берегу [1]), также связана с такими точными корреляциями (на этот раз в движении частиц вблизи камня на дне пруда). Таким образом, в этих ситуациях нам нет необходимости привлекать дополнительные гипотезы для объяснения запаздывающего распространения ряби. Гипотеза об энтропии уже достаточна для исключения двух ситуаций с опережением как совершенно невероятных, но она исключает также описанную выше бессмысленную ситуацию с запаздыванием, когда камень вылетает из воды (конечно, в предположении, что нет никакого иного фактора, ответственного за вылет камня, скажем пловца-подводника и т. п.).

В случае электромагнитного излучения ситуация в основном такая же, как с рябью. Небольшое отличие возникает, если рассматривать звезды, сияющие в практически пустой Вселенной; вполне возможно, что часть их излучения никогда не поглотится какой-либо материей, а будет существовать неопределенно долго, пока Вселенная расширяется или пока она не закончит свое существование в некоторой пространственно-временно́й сингулярности. Могло бы существовать также и излучение без источников, возникшее непосредственно в момент «большого взрыва» или в белой дыре или, возможно, даже пришедшее из бесконечности от ранее сколлапсировавшей фазы Вселенной.

Едва ли эти возможности в действительности способны что-либо существенно изменить в наших рассуждениях. Я упоминаю о них в основном потому, что так называемой «абсорбционной теории излучений» было посвящено много работ. По этой теории [42, 43], вклад в электромагнитное излучение от каждого заряда берется наполовину запаздывающим и наполовину опережающим, а присутствие дополнительно к этому излучения без источников или без стоков считается «нежелательным». Постулируя отсутствие такого дополнительного излучения, устанавливают связь между расширением Вселенной и запаздыванием излучения, хотя, на мой взгляд, и не очень убедительно. (Я вынужден признаться, что вся эта программа вызывает у меня не слишком много симпатий: она коробит меня своей предвзятостью к бедняге фотону, которому не позволяется иметь такие же степени свободы, как у всех массивных частиц!)

Во всяком случае, применимость принципа энтропии к вопросу о запаздывании, как мне кажется, совершенно не зависит от этого

---

[1]) Читатель может заметить близкую связь между этой ситуацией и движением типа «заг», описанным Голдом [9, 41]. Сходящиеся волны, встречающие падающий камень, относятся к типу «заг», тогда как два других случая относятся к типу «зиг».

[44]. Присутствие свободного излучения, приходящего из бесконечности (или, скажем, из сингулярности «большого взрыва») и сходящегося в прожекторе в момент его включения, или иной подобный абсурд в не меньшей мере соответствуют корреляциям того типа, который уменьшает энтропию в начальном состоянии, чем если бы это излучение приходило от источников. Единственное отличие состоит в том, что эти корреляции налагаются непосредственно на сами фотоны, а не на испускающие их частицы. Следовательно, нужно было бы ожидать, что таких корреляций не будет, если энтропия в будущем должна быть велика. Соответственно не может быть возражений против того, чтобы эти корреляции (в обращенной по времени форме) содержались в *будущих* граничных условиях, поскольку энтропия была мала в прошлом — а это, конечно, необходимо для того, чтобы звезды могли светить!

У читателя может возникнуть вопрос: как в действительности задавать граничные условия — на бесконечности или на пространственно-временнóй сингулярности, чтобы можно было подробно обсуждать такие корреляции? Конечно, здесь могут возникнуть серьезные технические трудности, особенно в случае сингулярностей. Но детали этих трудностей не должны существенно повлиять на предыдущие рассуждения, по крайней мере если предположение о космической цензуре [1]) является верным. Я хотел бы отложить эти вопросы до разд. 3.2, отметив лишь, что в определенных условиях (например, в такой модели «большого взрыва», в которой полный заряд внутри светового конуса прошлого некоторого наблюдателя отличен от нуля) возникает необходимость в определенном количестве излучения без источников (а в других случаях — в определенном количестве излучения без стоков) [45]. Нет никаких оснований считать, что это излучение должно быть коррелировано каким-либо образом, не согласующимся с принципом энтропии. Звезды попрежнему будут испускать свет, а не «вбирать» его независимо от того, имеется или нет какое-либо добавочное излучение, пронизывающее пространство, при том, однако, условии, что интенсивность этого излучения меньше интенсивности излучения звезд, а также при условии отсутствия специальных корреляций.

2.5. ПСИХОЛОГИЧЕСКОЕ ВРЕМЯ

По иронии природы труднее всего понять ту стрелу, которая самым непосредственным образом связана с нашим жизненным опытом, а именно то чувство неумолимого хода времени, в соответствии с которым потенциальное, как нам кажется, превращается в актуальное. Но с появлением специальной теории относительности стало ясно, что это ощущение, по крайней мере в *некоторых* отно-

[1]) Под космической цензурой подразумевается принцип, который запрещает присутствие наблюдаемых «голых» (не скрытых за горизонтом), или обнаженных, сингулярностей.— *Прим. перев.*

шениях, иллюзорно. Человеку свойственна инстинктивная (или приобретенная) убежденность в универсальности его собственного понятия о ходе времени, вследствие чего он считает, что превращение потенциального в актуальное (а каждый из нас чувствует, что такое превращение действительно имеет место) обязано происходить одновременно для всех нас. Специальная теория относительности говорит нам об ошибочности такого взгляда на мир (и именно этот вывод, вероятно, был самым большим препятствием к пониманию и принятию этой теории). Два человека трусят один за другим по улице. Какие потенциальные события превращаются в актуальные в то же время на какой-нибудь планете в туманности Андромеды? С точки зрения наших двух бегунов, могло бы быть расхождение в несколько дней! [1]) (Если же принять точку зрения, что события становятся актуальными, скажем, на чьем-либо прошлом или будущем световом конусе, а не по эйнштейновскому определению одновременности, от этого субъективность оценки происходящего только усиливается.)

Таким образом, *кажется*, что теория относительности ведет к картине, в которой утверждение «потенциальное превращается в актуальное» либо крайне субъективно, либо бессмысленно. Тем не менее в нас крепко держится ощущение, что между прошлым и будущим имеется самое фундаментальное различие, а именно: прошлое реально, и его нельзя изменить, тогда как на будущее можно влиять, и оно в некотором смысле не фиксировано. В представлении

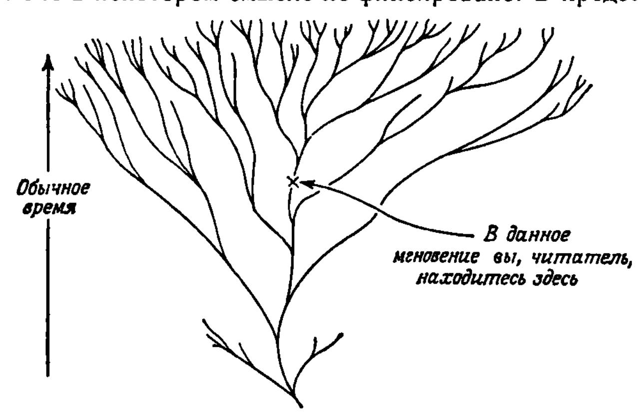


Рис. 2. Модельная вселенная, разветвляющаяся в будущее.

о мире, соответствующем обычной трактовке теории относительности, эта картина, естественно, отвергается, вместо нее вводится четырехмерное жестко детерминированное описание, и, таким обра-

[1]) Чтобы придать вопросу бóльшую определенность, вообразим, что «в этот самый момент» на этой планете заседает некий комитет, решающий будущее человечества!

зом, оказывается, что наши инстинктивные представления относительно возможности изменить будущее иллюзорны.

Однако я не думаю, что от наших представлений нужно полностью отказаться. Можно вообразить модельные вселенные, в которых будущее остается недетерминированным. Представим себе непрерывно ветвящуюся вселенную, подобную изображенной на рис. 2. Я могу представить себя находящимся в некоторой средней точке вселенной, которая «поднимается» вверх по рисунку по мере течение моего психологического времени. В этой модели разветвление происходит только в будущем. Путь в прошлое от этой точки (т. е. прошлая история вселенной) абсолютно однозначен, и в то же время имеется множество альтернативных путей в будущее (т. е. данному нынешнему состоянию соответствует много альтернативных возможных будущих историй вселенной).

Имеются (по крайней мере) два способа сделать такую модель релятивистской. По первому способу (рис. 3, *а*) разветвление происходит вдоль будущих световых конусов точек, в которых производится «наблюдение» (предположительно квантовомеханическое), по второму — вдоль прошлых световых конусов этих точек. Второй способ (рис. 3, *б*) представляется значительно менее правдоподобным, чем первый, поскольку вселенная должна будет «знать», что ей придется разветвиться раньше наблюдения, а это предполагает наличие некоторого рода точных корреляций в прошлом, подобных тем, которые ведут к излучению типа опережающего. В то же время первая мо-

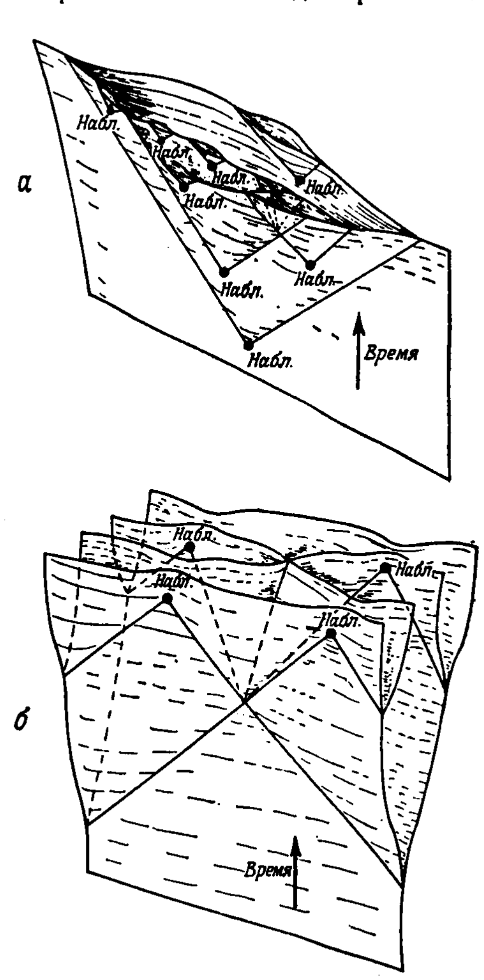


Рис. 3. Два возможных варианта релятивистской, разветвляющейся в будущее вселенной. «Набл». означает событие, в котором производится некоторое «наблюдение».

дель не совсем неправдоподобна. Можно представить себе, например, что разветвление сопровождается определенного рода запаздывающей редукцией волновой функции, причем на разных ветвях развитие волновой функции начинается с разных собственных векторов оператора, соответствующего наблюдению.

Такого рода модель можно назвать «вселенной типа Эверетта» [26—29], хотя для меня вовсе не ясно, является ли это разновидностью картины, к которой ведет квантовомеханический формализм Эверетта, и вот почему. Во-первых, если картина Эверетта является по существу переформулировкой обычной квантовой механики, она должна быть обратимой по времени, т. е. должно быть столько же разветвлений в прошлое, сколько в будущее. Во-вторых, в картине Эверетта рассматривается единственная волновая функция для всей вселенной, которая сама никогда не «редуцируется». Вместо этого она естественным образом представляется в виде линейной комбинации состояний, в каждом из которых имеются измеряющий прибор и физическая система, оказывающаяся в отдельном собственном состоянии прибора. Не исключено, что при дополнительных допущениях, скажем относительно природы рассматриваемого гамильтониана и т. п., можно доказать, что картина Эверетта в определенном приближении приводит к чему-то, напоминающему рис. 2 и 3. Но, по-видимому, этого никто не сделал. В частности, без каких-то сильно асимметричных по времени дополнительных предположений будет невозможно исключить обращения по времени рис. 2 и 3 или бесчисленные возможности, в которых разветвление происходит как в прошлое, так и в будущее, причем многие из ветвей расходятся лишь временно и затем снова сливаются.

Такие рассуждения, несомненно, в высшей степени спекулятивны. Однако при этом мы затрагиваем вопросы, которые вообще едва ли осознаются с точки зрения физики,— особенно те из них, где явно замешано человеческое «сознание» или нечто внечеловеческое; с этими вопросами тесно связана вся проблема психологического времени, а картина Эверетта может и не иметь к ним отношения. Пространственно-временные модели, напоминающие, скажем, рис. 3, *а*, наверное, заслуживают изучения. Они могли бы узаконить некоторые математические объекты как, например, лоренцевы многообразия, подчиняющиеся уравнениям Эйнштейна, но без условия хаусдорфовости [46]. Допустимо думать, что такая модель находится в большем соответствии с нашим интуитивным ощущением детерминированности прошлого и индетерминированности будущего, чем наше обычное представление о хаусдорфовом пространстве-времени. И мы могли бы сказать, что наше ощущение «время течет в будущее, а не в прошлое» является естественным ввиду того, что именно в направлении будущего потенциальное становится актуальным. Такую модель можно было бы рассматривать как «объективное» описание мира, содержащего определенные чисто «субъективные» элементы. Можно было бы вообразить также других наделенных

разумом наблюдателей, бредущих иными путями по мириадам ветвей (то ли по воле случая, то ли даже проявляя определенную «свободу воли»). И каждый такой наблюдатель имел бы свое «субъективное» понятие о мире.

Нельзя сказать, что я имею сколько-нибудь серьезную склонность верить в такую картину. Особое неудобство я вижу в той возможности, что мои друзья исчезнут в разных ветвях вселенной и вокруг не останется ни одного умного человека, с кем стоило бы поговорить. Так или иначе, в этих рассуждениях я слишком отклонился от моей провозглашенной ранее приверженности к общепринятому и при этом не пришел к чему-либо новому в вопросе о *происхождении* асимметрии по времени. Поэтому я должен твердо укрепиться в вере, троекратно повторив про себя: «Пространство-время есть *хаусдорфово* дифференцируемое многообразие; пространство-время есть *хаусдорфово*...»!

Пока еще я не делал никакой попытки связать *направление* психологического времени с ключевым вопросом об энтропии. Ясно, что любая такая попытка должна дать весьма неполный ответ ввиду нашего крайне смутного понимания того, что́ представляет собой психологическое время. При этом следует подчеркнуть, что данный вопрос не есть в точности вопрос большей (как это может показаться) *познаваемости* прошлого, чем будущего. В самом деле, не всегда верно, что прошлое легче установить достоверно, чем будущее, т. е. что «послесказание» (угадывание прошлого) более достоверно, чем «предсказание» — угадывание будущего. На конференции в Корнелле в 1963 г. Голд [41] заметил, что куда легче предсказать последующее поведение недавно запущенного спутника, чем точно установить время и место его запуска. Более того, не совсем ясно, почему феномен роста энтропии должен всегда сочетаться с простотой «послесказания» и трудностью предсказания. Если предшествующие состояния являются более «исключительными», а последующие более «обычными», то, казалось бы, мы стоим на более твердой почве при предсказании последующего состояния, чем при «послесказании» предшествующего. Поскольку все же на практике «послесказание» обычно легче (или по меньшей мере более точно), чем предсказание (воспоминания более надежны, чем прогнозы), точная связь между этим вопросом и вопросом об энтропии очень неясна.

Однако в действительности вопрос стоит несколько иначе. Важна не легкость суждения о прошлом, а ощущение невозможности изменить прошлое. Подобным же образом для нас существенны не те трудности, которые могут встретиться при попытках догадаться о будущем (или вычислить его), но ощущение, что на будущее мы можем влиять. Таким образом, наряду с замечанием Голда о точной предсказуемости будущего движения спутника могла бы возникнуть и тягостная тревога, что кто-то (обладающий всеми доступными в данное время техническими средствами) может попы-

таться сбить его. С другой стороны, кажется совершенно немыслимым, чтобы можно было предпринять какую-либо акцию, которая изменила бы дату запуска *после* того, как спутник оказался на орбите. Если же в принципе будущее не обязательно существенно отличается от прошлого, то мы должны допустить кошмарную возможность [47], что даже «фиксированное» прошлое можно было бы при определенных обстоятельствах изменить по нашему усмотрению. И не становится ли тогда единственным препятствием к этому проблема «денег»? (Поневоле приходит в голову идея из научной фантастики о всемогущем правительстве, поднимающемся на ступень выше простой фальсификации истории — на деле *изменяющем* прошлое!) Я предпочту обойти этот вопрос стороной.

Имеется, однако, более плодотворная точка зрения на интуитивное различие прошлого и будущего. Мы склонны относиться к событиям в прошлом как к «причинам», или «основаниям», для будущих событий, но не наоборот. Это по меньшей мере *совместимо* со взглядами, изложенными мной в разд. 2.3 и 2.4. Моя позиция состояла в том, что низкая энтропия в некоторый момент времени может рассматриваться как «причина» точных корреляций в движениях частиц в другой момент, но наличие или отсутствие таких корреляций нельзя рассматривать как «причину» для чего-нибудь еще. Мы наблюдаем, что энтропия была мала в прошлом, и ожидаем, что будут иметь место точные корреляции в будущем. Таким образом, именно малость энтропии в прошлом приводит к тому, что состояние в прошлом обеспечивает «причинами» состояние в будущем, но не наоборот. Это представляется мне вполне разумным. Если бы в нашем мире случалось так, что куча винтиков и пружинок иногда спонтанно собирались в идущие часы, то люди определенно не отказались бы приписать каким-то событиям в будущем, будто они являются «причинами» таких оказий. Эти происшествия могли бы сосуществовать с другими, более привычного сорта, «причинами» которых считались бы события в прошлом. Однако наша Вселенная устроена не совсем так! «Причины» хода событий в этих *обоих* типах вселенной прослеживаются к ситуациям с низкой энтропией, но в нашей Вселенной эти низкоэнтропийные состояния оказываются в прошлом.

Итак, по крайней мере в этом случае, наше психологическое ощущение разницы между прошлым и будущим может быть непосредственно сведено к вопросу об энтропии. Возможно, что такую связь можно обнаружить и в других аспектах.

2.6. РАСШИРЕНИЕ ВСЕЛЕННОЙ

В разд. 2.4 я косвенно указывал на то, что расширение Вселенной не может считаться непосредственной причиной запаздывания излучения; феномен запаздывания является всего лишь одним из многих следствий предположения, что первоначальное состояние Вселенной имело значительно меньшую энтропию, чем будет иметь

ее конечное состояние (и соответственно, что в этом начальном состоянии отсутствовали корреляции, понижающие энтропию). Теперь я хочу привести доводы в пользу того, что расширение Вселенной не может быть ответственно за эту разницу в энтропии.

В самом деле, предположим, что имеет место противное и что по какой-то причине рост энтропии непременно сопутствует расширению Вселенной. По симметрии относительно обращения времени такая точка зрения повлекла бы за собой утверждение, что в сжимающейся Вселенной энтропия должна убывать [48]. При этом следует рассмотреть две основные возможности. Во-первых, может оказаться, что расширение нашей реальной Вселенной в один прекрасный день сменится на сжатие и в этом случае, согласно нашему допущению, энтропия начнет убывать, устремляясь к низкому конечному значению. Вторая возможность состоит в том, что расширение будет продолжаться вечно — до достижения состояния максимальной энтропии (мы игнорируем вопрос о циклах Пуанкаре и т. п.).

По-видимому, имеются самые серьезные возражения против идеи, что тенденция к росту энтропии сменится на обратную, когда Вселенная достигнет максимума расширения. Трудно представить, как такое обращение может произойти без достижения некоторого рода теплового равновесия на полпути. Иначе, думается, можно было бы представить себе где-то посередине состояние, в котором явления обычного типа (например, запаздывающее излучение и разбивающиеся часы) сосуществовали бы с явлениями типа обращенных во времени (например, с опережающим излучением и самовосстанавливающимися часами). Рассуждать о подобных ситуациях «в целях аргументации», конечно, можно, но допускать их всерьез для нашей *реальной* Вселенной — это все-таки совсем другое дело. Более того, в момент достижения симметрии по времени, по-видимому, все еще приходило бы запаздывающее излучение очень далеких галактик (эти удаленные галактики все еще были бы видны как разбегающиеся) и в то же время имелось бы опережающее излучение с обращенным во времени поведением (т. е. специальным образом коррелированное и сходящееся на слетающихся галактиках). Тут возникли бы серьезные проблемы самосогласованности [9] (хотя я не берусь утверждать, что они абсолютно непреодолимы). Я лично не нахожу возможности воспринимать всерьез подобную картину, но, видимо, не у всех интуиция оказывается столь скованной [48]!

Мы могли бы, наоборот, предположить, что масштабы времени для обращения расширения столь велики, что к моменту максимального расширения достигается эффективное тепловое равновесие. Но времена, которые приходится допускать для этого, совсем другого порядка, чем обычные космологические масштабы. *По существу*, такая вселенная как бы вообще не переходит к сжатию, и ситуация должна рассматриваться наравне с моделями бесконечно расширяющихся вселенных.

Может возникнуть мысль, что в этих последних моделях удается избежать рассматриваемых здесь проблем, но это неверно. Представим себе астронавта в такого рода вселенной, падающего в черную

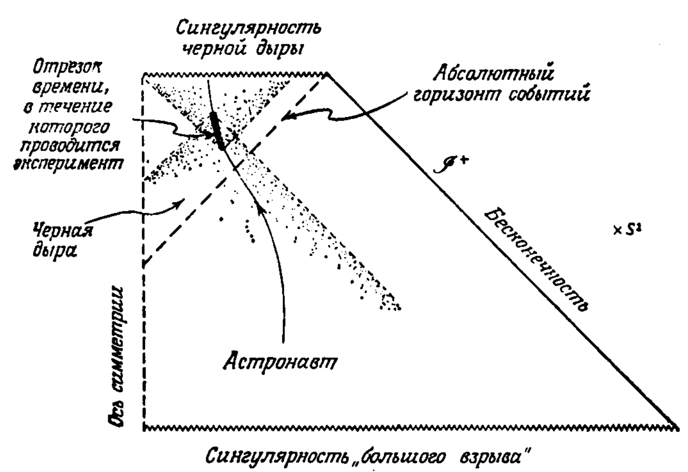


Рис. 4. Конформная диаграмма астронавта, падающего в черную дыру во фридмановской вселенной ($k=-1$).

дыру. Для определенности предположим, что масса дыры $10^{10}M_\odot$, так что в распоряжении астронавта внутри дыры будет около суток, причем бо́льшую часть этого времени он не будет испытывать действия значительных приливных сил и сможет не спеша провести эксперименты. Вся эта ситуация изображена на рис. 4 в виде стандартной конформной диаграммы (световые конусы проходят под углом 45°, и предполагается наличие сферической симметрии). Заметим, что вся история астронавта за точкой пересечения им абсолютного горизонта событий лежит внутри прошлой области зависимости для сингулярности черной дыры, а также внутри будущей области зависимости сингулярности «большого взрыва». Допустим, что, находясь внутри дыры, астронавт в течение некоторого времени проводит эксперименты. Поведение его аппаратуры (и, конечно, жизненные процессы в его собственном теле тоже) полностью заданы условиями у сингулярности черной дыры (в предположении, что это поведение подчиняется обычным дифференциальным уравнениям гиперболического типа), но равным образом они полностью определяются и условиями у сингулярности «большого взрыва». Ситуация внутри черной дыры по существу не отличается от поздних стадий реколлапсирующей вселенной. Если ваша точка зрения сводится к тому, что локальная стрела времени непосредственно связана с расширением вселенной, то вы, несомненно, должны будете рассчитывать, что в ходе экспериментов нашего астронавта энтропия непременно будет *убывать* (по отношению к «нормальному»

времени). В итоге вы, по-видимому, должны будете прийти к мысли, что астронавт будет уверен в том, что он вылетает из дыры, а не падает в нее (предполагается, что жизненные процессы в его организме не нарушатся при столь крутой смене нормального процесса нарастания энтропии на противоположный).

Я вынужден заявить, что такую возможность могу рассматривать только как еще *более* абсурдную, чем даже обращение энтропии при максимальном расширении реколлапсирующей вселенной! Возможно, не следует предполагать, что обращение энтропии произойдет внезапно, как только астронавт пересечет горизонт (и только с этого момента коллапсирующий аспект его граничных условий *решительно* перевесит аспект расширения). В таком случае странное поведение энтропии должно было бы проявиться задолго до реального пересечения астронавтом горизонта, а тогда он мог бы передумать: включить ускорение в сторону от дыры и, избежав захвата ею, получить возможность рассказать о своих удивительных открытиях остальному миру! Это рассуждение с черной дырой приложимо в действительности и к реколлапсирующей вселенной. Нет необходимости ждать реколлапса всей вселенной, чтобы нелепости этой точки зрения начали себя обнаруживать. (Может оказаться, что некоторые приверженцы этой точки зрения вообще не склонны признавать реальность черных дыр. У меня нет желания приводить здесь доводы, на мой взгляд весьма убедительные, в пользу существования черных дыр, и я сошлюсь только на работы [7, 51].)

Можно было бы выдвинуть аргумент, что пространство-время, изображенное на рис. 4, основано на слишком сильном и нереалистическом предположении о сферической симметрии. На самом деле это не так: отказ от сферической симметрии не внесет качественных отличий в нашу картину при условии, что делается (достаточно сильное) предположение о космической цензуре. Я предпочту отложить более подробное обсуждение этого вопроса до разд. 3.2. Сейчас же мне достаточно того, что я снова могу вернуться к общепринятому взгляду и сделать вывод, что расширение вселенной *само по себе не несет* какой-либо ответственности за факт возрастания энтропии в нашей Вселенной.

Я вовсе не хочу сказать, что рассматриваю соответствие между этими двумя важнейшими фактами как совершенно случайное. Отнюдь нет. Позднее я как раз и попытаюсь доказать, что оба они являются следствиями особого свойства «большого взрыва» — свойства, какого *не следует* ожидать у сингулярности реколлапса.

2.7. ЧЕРНЫЕ ИЛИ БЕЛЫЕ ДЫРЫ?

Общая теория относительности является теорией, симметричной по времени. Поэтому каждому решению ее уравнений (с симметричными по времени уравнениями состояния), асимметричному по вре-

мени, должно соответствовать другое решение, в котором имеет место обратное упорядочение по времени [1]). Одним из наиболее известных решений является то, которое соответствует (сферически-симметричному) коллапсу звезды (описываемой, скажем, с помощью тензора энергии-импульса T_{ab} для «пыли»), которая превращается

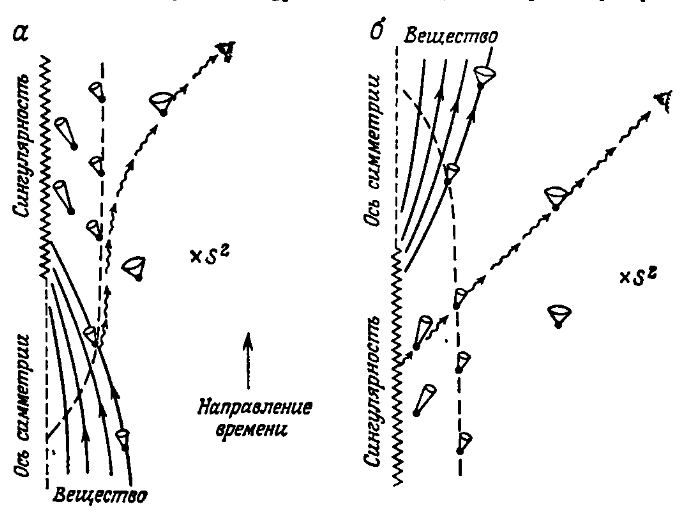


Рис. 5. Черные и белые дыры (диаграммы Финкельштейна).
а — коллапс в черную дыру; б — взрыв белой дыры Наблюдатель смотрит справа сверху.

при этом в черную дыру [7, 54, 55]. Обращение по времени этого процесса и есть то, что называют «белой дырой»; она в конечном счете взрывается, порождая облако материи. Пространственно-временные диаграммы этих двух процессов приведены на рис. 5.

Разные авторы [52, 53] пытались привлечь белые дыры для объяснения квазаров и других необычных астрономических явлений. Напомню, что при (классическом) коллапсе в черную дыру процесс стартует от совершенно обычного распределения материи, подчиненной детерминистическим эволюционным уравнениям. На определенной стадии может образоваться ловушечная поверхность, ведущая к существованию *абсолютного горизонта событий*, за который частицы могут упасть, но ни одна не может вырваться обратно. После того как вся имеющаяся материя будет поглощена, устанавливается дыра, которая остается неизменной до скончания времени (этим концом может стать реколлапс вселенной). (Здесь игнорируются квантовомеханические процессы Хокинга [56], которые мы вскоре обсудим.) Следовательно, белая дыра возникает в начале

[1]) Для этого «решение уравнений Эйнштейна» должно быть четырехмерным лоренцевым многообразием, *ориентированным по времени* (и возможно, пространственно-ориентированным). «Обращение упорядочения по времени» сводится к выбору противоположной ориентации времени.

времени (т. е. при «большом взрыве») и остается в практически неизменном состоянии в течение неопределенно долгого времени. Затем она исчезает, разлетаясь со взрывом в облако обычной материи. В течение продолжительного спокойного периода границей белой дыры служит некоторый стационарный горизонт, а именно *абсолютный горизонт частиц*, за который ни одна частица не может залететь, но из-за которого в принципе могут вылетать частицы.

В этом предполагаемом поведении белой дыры есть нечто, кажущееся «термодинамически довольно неудовлетворительным» (или физически довольно маловероятным), хотя и трудно с определенностью уловить, что́ именно представляется неверным. Обычная картина коллапса в черную дыру кажется «удовлетворяющей» нашим общепринятым представлениям о классическом детерминизме. Если предположить, что имеет место строгая космическая цензура [57—59], то все пространство-время детерминировано в *будущее* от некоторой «подходящей» гиперповерхности Коши, кривизны на которой малы. В случае же белой дыры нет возможности задания таких граничных условий в прошлом, поскольку начальная гиперповерхность Коши должна пересечься с сингулярностью (или подойти к ней слишком близко). Иными словами, будущее поведение такой белой дыры, по-видимому, не определяется сколько-нибудь *заметным* образом ее прошлым. В частности, момент времени, в котором белая дыра взорвется и превратится в обычную материю, по-видимому, «выбирается» ею самой, и его нельзя точно предсказать с помощью обычных физических законов. Конечно, можно было бы использовать граничные условия в будущем для послесказания поведения белой дыры, но в нашей Вселенной, где энтропия растет, это — термодинамически неестественный обходной путь. (И уж во всяком случае можно просто обратиться к памяти как к более эффективному способу послесказания!)

С этой неопределенностью связан тот факт, что в случае белой дыры наблюдатель (использующий обычный запаздывающий свет) может непосредственно видеть сингулярность, в то время как в случае черной дыры он этого сделать не может (рис. 5). Поскольку предполагается, что пространственно-временна́я сингулярность — это место, где известные законы физики нарушаются, то может быть и не удивительно, что в поведении белой дыры возникает столь сильный элемент индетерминизма. В таком случае сингулярности могут оказывать причинное воздействие на внешний мир (вне горизонта).

Согласно принятым сейчас представлениям о физических процессах, которые сопутствуют сингулярностям, ожидается, что должно происходить *рождение частиц* [62, 63, 66]. Это общий вывод ряда различных исследований по квантовой теории поля в искривленном пространстве-времени. Однако ввиду незавершенного состояния этой теории такие исследования не всегда согласуются в деталях выводов. В приложении этой теории к белым дырам в настоящее

время выделились два идейно разных направления. Согдасно Зельдовичу [64], белая дыра полностью неустойчива по отношению к указанному процессу рождения и должна испариться мгновенно, тогда как Хокинг [67] выдвинул остроумную идею, что процесс испарения белой дыры должен происходить значительно медленнее и по своей природе должен быть неотличим от испарения черной дыры той же массы (процесс Хокинга) или, попросту говоря, что белая дыра должна быть неотличима от черной дыры! Эта точка зрения Хокинга, будучи во многих отношениях весьма радикальной, несет с собой ряд серьезных трудностей, которые мы вскоре рассмотрим. Первая же из этих двух точек зрения ведет к утверждению, что белые звезды физически не существуют (хотя ввиду оценочного характера вычислений по рождению частиц им нельзя придавать слишком большого значения; см., однако, также работу Эрдли [65]).

Имеется еще один довод, чтобы считать белые дыры антитермодинамическими объектами (хотя и этот довод должен быть видоизменен, если пытаться охватить вышеупомянутые более радикальные идеи Хокинга). Согласно формуле Бекенштейна — Хокинга [56, 69], площадь поверхности A абсолютного горизонта событий *черной* дыры пропорциональна S, собственной энтропии дыры:

$$S = kAc^3/4\hbar G$$

(k — постоянная Больцмана, G — ньютоновская гравитационная постоянная). Принцип площадей классической общей теории относительности [7, 70, 71] утверждает, что в классических процессах A не убывает со временем, а это согласуется с термодинамической стрелой времени, т. е. с тем, что энтропия не должна убывать. Если теперь мы должны подобным же образом приписать собственную энтропию белой дыре, то трудно понять, каким образом значение этой энтропии могло бы отличаться от того, которое дает та же формула Бекенштейна — Хокинга, где A означает теперь площадь абсолютного горизонта *частиц*. Обращенный во времени принцип площадей утверждает тогда, что в классических процессах величина A *не возрастает*, т. е. все происходит в направлении, обратном обычной термодинамической стреле времени для энтропии. В частности, значение A должно существенно *уменьшаться* всякий раз, как только из белой дыры выплескивается заметное количество материи, как, например, при финальном взрыве, изображенном на рис. 5. Таким образом, имеет место сугубо антитермодинамическое поведение.

Итак, что касается физики белых дыр, то здесь, по-видимому, имеются две основные возможности, заслуживающие рассмотрения. Одна из них состоит в том, что должен быть какой-то общий принцип, который исключает их существование (или по крайней мере делает его крайне маловероятным). Другая возможность содержится в вышеупомянутом ходе рассуждений Хокинга, из которого сле-

дует, что вследствие квантовомеханических эффектов черные и белые дыры нужно считать физически неразличимыми [67]. Обсудим сначала эту замечательную идею Хокинга. Затем я попытаюсь объяснить, почему я думаю, что это все же не может быть правильным объяснением, и считаю необходимым, чтобы белые дыры физически *не существовали*.

Напомню, во-первых, что́ представляет собой хокинговское излучение, которое, согласно вычислениям, сопутствует всякой черной дыре. Температура этого излучения обратно пропорциональна массе дыры, составляя обычно около 10^{-7} K для черной дыры массой $1M_{\odot}$. Таким образом, в случае дыр с массами порядка звездных температура излучения ничтожна, но она может играть заметную роль в случае очень малых дыр, если такие существуют. Если рассматривать остальную вселенную как пустую, то хокинговское излучение должно привести к тому, что черная дыра будет терять массу, становиться горячее, излучать еще больше, терять массу еще больше и т. д., так что весь процесс будет ускоряться, надо полагать, до тех пор, пока дыра не исчезнет в финальном взрыве. Но для черной дыры солнечной массы (или более) на этот процесс ушло бы больше 10^{53} времен Хаббла! Кроме того, если бы даже такой процесс мог начаться, пришлось бы ждать 10^{7} или около того времен Хаббла, прежде чем расширение вселенной привело к снижению нынешнего уровня фонового излучения ниже уровня излучения дыры (предполагается модель бесконечно расширяющейся вселенной)!

Несмотря на абсурдность подобных цифр, представляет определенных теоретический интерес составить умозрительную картину, как это сделал Хокинг [67, 72], того состояния теплового равновесия, к которому могла бы прийти черная дыра в большой полости с абсолютно отражающими стенками. Если эта полость достаточно велика для данного массово-энергетического содержимого (случай *а* на рис. 6), черная дыра, по-видимому, высветится полностью, предварительно поглотив всякое имевшееся в полости блуждающее вещество и не оставив после себя в конечном счете ничего, кроме теплового излучения (и, возможно, небольшого числа термализованных частиц). Это конечное состояние будет состоянием «теплового равновесия» с максимальной энтропией.

Если полость будет значительно меньше (случай *в* на рис. 6) или ее массово-энергетическое содержимое значительно больше (хотя и не настолько больше, чтобы вызвать коллапс всей полости), то достижимым состоянием с максимальной энтропии будет единственная сферическая черная дыра, находящаяся в тепловом равновесии с окружающим излучением. Устойчивость достигается здесь по следующей причине: если вследствие флуктуации дыра излучит чуть больше и соответственно подогреется, ее окружение при этом подогреется еще в большей степени, отдаст дыре больше, чем та излучила, и, таким образом, вернет ее к прежним размерам;

если же при флуктуации дыра излучит меньше, чем поглотит, то ее окружение охладится больше, чем дыра, и она снова вернется к равновесию (рис. 6, *в*).

Имеется также промежуточный случай *б*, в котором полость имеет промежуточные размеры по отношению к данному массово-энергетическому содержимому; черная дыра остается еще стабильной, но это соответствует лишь *локальному* максимуму энтропии, а абсолютный максимум приходится на состояние, в котором присутствует лишь тепловое излучение (и, возможно, термализованные частицы), а черной дыры нет. В этом случае черная дыра может

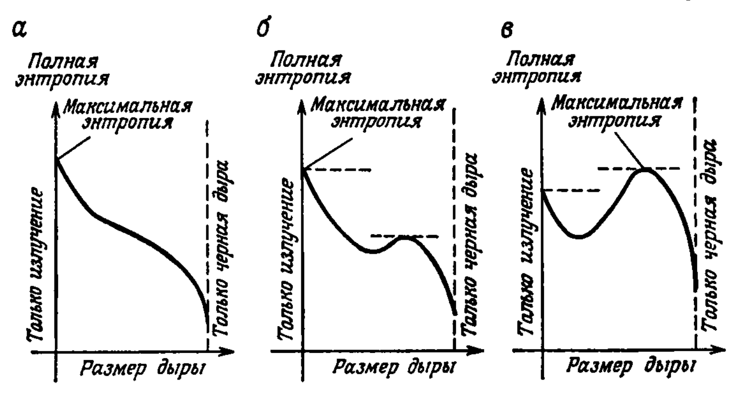


Рис. 6. Черная дыра Хокинга в полости с полностью отражающими стенками. *а* — большая полость; *б* — полость промежуточных размеров; *в* — относительно малая полость.

оставаться в равновесии с окружающим излучением в течение весьма длительного периода. Понадобилась бы огромная флуктуация, при которой было бы испущено значительное количество излучения, чтобы преодолеть низкоэнтропийный барьер между двумя локальными максимумами (рис. 6, *б*). При столь большой утечке массы из дыры она может разогреться значительно сильнее, чем окружение; тогда она потеряет еще больше массы, еще больше нагреется и, как в случае *а*, полностью высветится, что и приведет к нужному состоянию теплового излучения (плюс случайные термализованные частицы).

Следует подчеркнуть, что хотя мы и имеем здесь дело с процессами невообразимо больших временны́х масштабов [1]), эти ситуации имеют в определенном смысле фундаментальное значение для физики. Фактически здесь мы рассматриваем состояния с максимальной энтропией для *всех* физических процессов. В случаях *а* и *б* состояние с максимальной энтропией — это знакомая нам «тепловая

[1]) Если мы допустим существование (виртуальных) черных дыр размерами порядка 10^{-33} см (т. е. $\sim 10^{-20}$ радиуса элементарной частицы), то получим картину [73], для которой эти временны́е масштабы могут быть очень малы.

смерть вселенной», но в случае *в* мы встречаем нечто новое: черная дыра в тепловом равновесии с излучением. Конечно, с теоретическими деталями такой структуры сопряжено немало трудностей (например, броуновское движение черной дыры приводило бы к случайным столкновениям ее со стенкой полости, что привело бы к разрушению стенки). Такие проблемы здесь будут оставлены без внимания как не относящиеся к числу основных. Но все же, поскольку времена релаксации весьма велики по сравнению с нынешним возрастом вселенной, при интерпретации тех или иных выводов требуется определенная осторожность. Тем не менее думается, что здесь проясняются важные вещи.

Продолжая рассматривать доводы Хокинга, обратимся к случаю *в*. Бóльшую часть времени система остается вблизи максимума энтропии: она состоит из черной дыры и излучения. Однако вследствие большой начальной флуктуации, при которой дыра излучит значительное количество энергии, может начаться развитие событий, подобное случаю *б*, когда черная дыра испаряется, оставляя тепловое излучение. Но затем вслед за новым долгим периодом ожидания в достаточно малой области снова собирается (и снова вследствие флуктуации) излучение в количестве, достаточном для формирования черной дыры. Если эта дыра будет достаточно велика, система вновь вернется к состоянию с максимальной энтропией, в котором останется на весьма длительный срок.

Циклы, подобные этому, могут иметь место также в случае *б* (и даже в случае *а*), но с той разницей, что в них бóльшая часть времени приходится на состояние, в котором нет черной дыры, тогда как в случае *в* почти все время в системе присутствует черная дыра. Далее Хокинг утверждает, что равновесные состояния должны быть симметричными по времени, поскольку симметричны по времени все имеющие существенное отношение к делу физические теории (общая теория относительности, теория Максвелла, теория нейтрино, возможно, теории электронов, пионов и т. д., а также общий аппарат квантовой механики). Но в данном случае обращение времени ведет к белым дырам, а не к черным. Отсюда Хокинг приходит к выводу, что белые дыры должны быть физически неотличимы от черных дыр!

Это отождествление не столь абсурдно, как может показаться сначала. Хокинговское излучение от черной дыры оказывается переинтерпретированным как рождение частиц вблизи сингулярности белой дыры (и, таким образом, Хокинг предполагает довольно малую скорость рождения частиц вблизи сингулярности белой дыры). Исчезновение излучения в черной дыре становится обращенным во времени хокинговским излучением от белой дыры. Можно, конечно, представить себе падение в черную дыру некоторого сложного объекта, скажем телевизора. Неужели и это можно мыслить как обращенное во времени хокинговское излучение? Ответ таков: в процессе хокинговского излучения, поскольку оно тепловое [67, 74],

с равной вероятностью порождаются все возможные конфигурации. *Возможно*, чтобы часть хокинговского излучения представляла собой телевизор, но такое происшествие невообразимо маловероятно и соответствовало бы чудовищному уменьшению энтропии. Падение телевизора в черную дыру только кажется более «естественным», поскольку мы обращаемся здесь к ситуации, в которой энтропия в начальном состоянии мала. С равным успехом мы могли бы представить себе начальные граничные условия с низкой энтропией для обращенного по времени процесса хокинговского излучения, и именно это соответствовало бы аннигиляции телевизора при обращении по времени хокинговского излучения белой дыры.

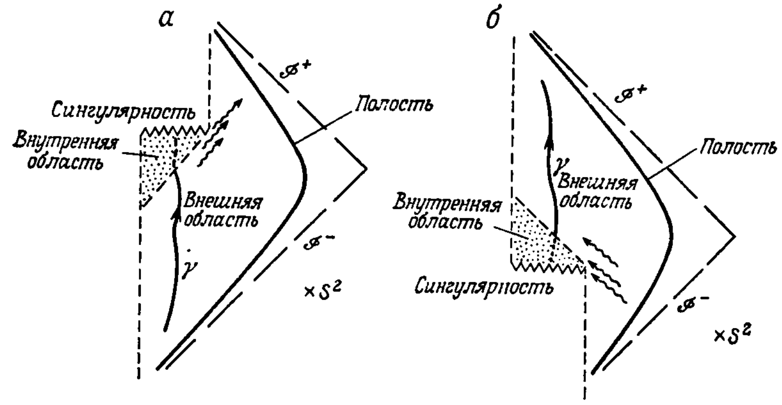


Рис. 7. Конформные диаграммы, иллюстрирующие асимметрию по времени появляющейся и исчезающей черной дыры.

а — классический коллапс с образованием черной дыры и последующим ее полным хокинговским испарением; *б* — хокинговская конденсация на белую дыру с последующим классическим исчезновением белой дыры.

Таким образом, пока что все это выглядит вполне приемлемым и вся схема в целом обнаруживает неожиданную элегантность и экономность. Но, к несчастью, она страдает двумя (а возможно, и тремя) серьезными пороками, которые, на мой взгляд, исключают ее из числа серьезных возможностей.

Во-первых, хотя геометрия пространства-времени снаружи горизонта стационарной черной дыры идентична геометрии снаружи горизонта стационарной белой дыры, это определенно *не так* для внешней геометрии черной дыры, возникающей при обычном гравитационном коллапсе, а затем исчезающей в соответствии с симметрией по времени описанного выше процесса Хокинга. Эта асимметрия по времени становится особенно ясной при использовании конформных диаграмм, которые приведены на рис. 7. Различия между внешними геометриями появляющихся и исчезающих черных и белых дыр можно точно описать в терминах их структуры множеств ГНП и ГНБ (см. разд. 3.2). Однако интуитивно различие

проявляется в существовании временииподобных кривых γ, которые в случае черной дыры «покидают» внешнюю область и «входят» в дыру, а в случае белой дыры — наоборот. Причина этого различия в том, что процесс *классического* коллапса не есть обращенный по времени *квантовый* процесс Хокинга. Нам не следует этому сильно удивляться, поскольку за эти два процесса ответственны совершенно разные физические теории (классическая общая теория относительности, с одной стороны, и квантовая теория поля на заданном фоне искривленного пространства-времени — с другой).

Точка зрения, которой придерживается Хокинг и при которой можно было бы избежать этих трудностей, состоит в том, чтобы рассматривать пространственно-временну́ю геометрию как в некоторой мере зависящую от наблюдателя. Коль скоро уж квантовая механика и геометрия искривленного пространства-времени оказались столь существенно переплетенными, утверждается с этой точки зрения, мы не можем последовательно говорить о классически объективном пространственно-временно́м многообразии. Наблюдатель, который падает в белую дыру, чтобы затем испариться в обращенном по времени хокинговском процессе, был бы, следовательно, уверен, что имеет дело с геометрией черной дыры, горизонт которой он пересекает и внутри которой его ожидает «классическая» участь быть уничтоженным в конце концов чрезмерными приливными силами.

Должен сказать, что принять эту картину мне почти столь же трудно, как и те, согласно которым энтропия начинает убывать, когда вселенная наблюдателя начинает вокруг него коллапсировать. Если бы рассматривались белые или черные дыры с радиусами порядка планковской длины ($\sim 10^{-33}$ см) или пусть даже порядка размеров элементарной частицы ($\sim 10^{-13}$ см), тогда такая неопределенность в геометрии еще была бы приемлема. Но для черной дыры с массой порядка солнечной (или больше) это потребовало бы весьма радикального пересмотра наших взглядов на геометрию, такого, который сильно сказался бы на всех приложениях общей теории относительности к астрофизическим явлениям. Правда, в разд. 2.5 кратко рассмотрен вариант картины мира, который допускает элемент «зависимости» геометрии от наблюдателя. Но все же я пока не вижу никакого способа связать эту точку зрения с неопределенностью в классической геометрии такого типа, к которому, по-видимому, приводит физическое отождествление черных дыр с белыми дырами.

Есть также и другие возражения против попытки рассматривать классический гравитационный коллапс как обращение по времени квантовомеханического процесса рождения частиц. Одно из них направлено не против самого отождествления процесса Хокинга с обращением по времени классического «проваливания» материи в черную дыру, а скорее против отождествления этих процессов с рождением частиц в областях с большой пространственно-времен-

нóй кривизной. Такое дополнительное отождествление не следует прямо из обсуждавшейся до сих пор симметричной по времени картины: вполне может оказаться, что происходящие явления имеют какой-то другой характер. Обратимся снова к представлению о хокинговском излучении черной дыры как о процессе рождения частиц вблизи сингулярности белой дыры. Если такое представление и в самом деле допустимо, то здесь мы имеем дело вовсе не с «нормальным» процессом рождения частиц в областях с большой кривизной, который неоднократно обсуждался в литературе [62], ибо в «нормальном» процессе частицы всегда рождаются *пáрами:* барион с антибарионом, лептон с антилептоном, положительно заряженная частица с отрицательно заряженной. Процесс же Хокинга явно не такой природы, о чем говорит его тепловой характер (для частиц, уходящих на бесконечность) [67, 74, 75].

Различие становится еще более явным, если мы пытаемся связать это попарное рождение частиц вблизи сингулярности белой дыры с обращением по времени процесса аннигиляции вещества вблизи сингулярности черной дыры. В самом деле, нет никаких ограничений на форму материи, которую могла бы классически поглотить черная дыра, и если в классических процессах выполняется строгая космическая цензура, то, по-видимому, даже индивидуальные заряженные частицы должны по отдельности разрушаться у сингулярности. (Этот пункт будет усилен в разд. 3.2.) Невозможно предположить, что частицы каким-то образом ухитрятся рассортироваться на пары «частица — античастица», прежде чем столкнутся с сингулярностью. Возможно, трудность здесь заключается не только и не столько в хокинговском отождествлении черных и белых дыр, сколько в самом подходе, допускающем, что процессы исчезновения и возникновения материи могут быть описаны в рамках известной нам физики. Таким образом, я утверждаю следующее: в то время как *допустимо* думать, что возникновение материи в «большом взрыве» можно трактовать в рамках известных (или хотя бы частично понятых) процессов рождения частиц, по-видимому, такое же допущение в отношении процессов разрушения вблизи сингулярности черной дыры, а если Хокинг прав, то и в отношении процессов рождения вблизи сингулярности белой дыры будет *неверным*. Следовательно, точка зрения Хокинга, по всей видимости, ведет к прямому противоречию с часто выражаемой надеждой, что возникновение частиц при «большом взрыве» может быть понято через процессы рождения частиц пространственно-временнóй кривизной. Это имеет прямое отношение к вопросу, насколько симметричная по времени физика может сохранить силу вблизи пространственно-временны́х сингулярностей. Это ключевой вопрос, и мы рассмотрим его подробнее в разд. 3.

В еще более прямой связи со схемой Хокинга находится трудность, возникающая при детальном изучении тех циклов, в которых, согласно рассмотренным выше случаям *б* и *в*, черная дыра

с массой порядка солнечной, находясь в устойчивом равновесии с окружением в нашей идеально отражающей полости, может исчезнуть и вновь появиться вследствие флуктуаций. Как в действительности выглядит наиболее вероятный путь к полному испарению черной дыры? Она могла бы, конечно, попросту сбросить всю свою массу в одной гигантской флуктуации. Но со случайным процессом Хокинга это происходило бы немыслимо редко. Несравненно больше вероятность высвечивания в одной гигантской флуктуации такой *части* массы дыры, которая нужна для того, чтобы поднять хокинг-бекенштейновскую температуру значительно выше той, до которой при этом поднимется температура среды. После этого испарение будет продолжаться «нормальным» образом, и новые невероятные происшествия не будут нужны. А теперь рассмотрим окончательный взрыв, при котором дыра «нормально» исчезает. Появятся электроны и позитроны, за ними пионы, затем в последний момент возникнет все множество нестабильных частиц, которые будут сложным образом распадаться. Наконец, надо думать, из точки аннигиляции будут по отдельности вылетать протоны и антипротоны. Только много позднее, случайно сталкиваясь в хаотическом движении в полости, протоны и антипротоны постепенно будут взаимно аннигилировать (или, возможно, иногда распадаться в процессе Пати — Салама [76], скажем, на электроны и позитроны, большинство которых опять же аннигилирует).

Как же теперь будет выглядеть наиболее вероятный способ формирования заново черной дыры в полости, когда она подходит к другой точке устойчивого равновесия с окружением? Без сомнения, это *не может* быть обращением по времени только что описанного процесса исчезновения, в соответствии с которым сначала протоны и антипротоны должны были бы ухитриться (после длительной подготовки) возникнуть (без всякой необходимости) из фонового излучения, чтобы затем с невероятной точностью встретиться в некоторой ничтожно малой области с целью последовать (какая необходимость?!) некоторой весьма изощренной физике частиц и в результате встречи с точно направленными γ-лучами и т. п. образовать разнообразные нестабильные частицы и т. д. и т. п.; потом (также по хореографии наивысшей точности) новые частицы (пионы, затем электроны и позитроны) должны направиться туда же, также предварительно возникнув из фона в нужный момент и в нужной пропорции. Лишь после этого и само фоновое излучение падает туда же, создавая ту огромную массу, которая нужна для возникновения дыры.

Для меня важно не то, что это курьезное начало необходимо признать крайне невероятной частью процесса образования дыры заново. Могу представить себе, что против этого что-то можно возразить. Но такое начало не является необходимым. Процесс формирования *черной* дыры должен вступить в существенную стадию, когда излучение собирает само себя в достаточно малой области,

чтобы произошло то, что является фактически обычным гравитационным коллапсом. В действительности же крошечное ядро, которое образовалось бы при столь тщательной подготовке, по-видимому, как-то *воспрепятствовало* бы в силу своей чрезмерной температуры дальнейшему коллапсу.

Так что ж здесь неверно? Что за асимметричная по времени физика проникла контрабандой в описание «наиболее вероятной» моды исчезновения дыры, из-за которой между этой модой и обращением по времени «наиболее вероятной» моды нового появления дыры оказалось такое расхождение? Возможно, нет никакой контрабанды, если белые дыры в принципе существуют и просто являются объектами, *отличными* от черных дыр. Вышеописанные тщательные приготовления, не будучи необходимыми для черных дыр, могли бы потребоваться для белых дыр. В конце концов, ведь нужен какой-то изощренный механизм возникновения сингулярности белой дыры, которая, как видно из рис. 7, имеет структуру, совершенно отличную от структуры черной дыры. Может оказаться, что образование такой сингулярности представляет собой необычайно тонкий процесс, требующий, чтобы частицы нужного сорта и в точно определенном порядке были направлены в одну точку с высокой энергией и исключительной точностью. Возникает дополнительная кажущаяся трудность, связанная с тем, что мы должны высказать суждение о некоторой области пространства-времени (а именно области за горизонтом), которая не лежит в области зависимости какой-либо начальной гиперповерхности Коши, проведенной до появления белой дыры (рис. 7, *б*). Конечно, при (хокинговском) исчезновении черной дыры тоже возникает эта проблема (обращенная по времени), но как-то не принято заниматься «послесказанием» с гиперповерхности Коши, и потому это вызывает куда меньше беспокойства. Если же мы допускаем существование белых дыр, возникает еще одна трудность: мы сталкиваемся с упомянутой ранее проблемой предсказания того, чтó именно и когда белая дыра собирается излучить. Как я уже отмечал, *если* вблизи сингулярностей правомочна физика, симметричная по времени, то «обычные» идеи относительно рождения частиц из-за кривизны не работают. Концепция «хаотичности» Хокинга [67], может быть, и ближе к истине, но, к сожалению, теперь, когда отброшена важнейшая идея отождествления черных и белых дыр, она слишком неопределенна для того, чтобы, основываясь на ней, вычислять что-либо.

Я лично нахожу эту картину равновесия, включающую спорадическое появление таких белых дыр, крайне малоприятной. А ведь наверняка можно было бы придумывать и иные чудовищные зебро-подобные комбинации белого и черного! Думается, что все это не имеет никакого отношения к реальной физике (по крайней мере на макроскопическом уровне). Единственная цель, которая *вообще* вынуждала нас рассматривать белые дыры,— это спасение симметрии по времени! Возникающие при этом непредсказуемости и не-

приятности представляются слишком высокой ценой за нечто такое, что *даже и не выполняется* в нашей Вселенной при больших масштабах.

Одним из следствий гипотезы, которая выдвинута мною в следующем разделе, будет исключение сингулярности белой дыры как неприемлемого граничного условия. Эта гипотеза асимметрична по времени, но именно это и необходимо для объяснения других стрел времени. Добавив эту гипотезу к нашему рассмотрению равновесия внутри абсолютно отражающей полости, мы явственно увидим, какая асимметричная физика проникла сюда «контрабандой». Моя гипотеза сформулирована так, чтобы никоим образом не ограничивать поведение черных дыр, но она запрещает белые дыры и тем самым отбрасывает как ненужный тот фантастический сценарий, который, как мы видим, необходим для возникновения белых дыр.

Надеюсь, читатель простит мне столь пространные рассуждения о белых дырах, проведенные только для того, чтобы в конце провозгласить, что они не существуют! Однако гипотетические ситуации часто могут привести к пониманию важных вещей, особенно когда они граничат с парадоксальным, как это было здесь.

3. СИНГУЛЯРНОСТИ — КЛЮЧ К ПРОБЛЕМЕ?

Каков итог нашего обсуждения в данный момент? Согласно разд. 2.3—2.5, и стрела энтропии, и стрела запаздывающего излучения, и, *возможно*, стрела психологического времени могут быть объяснены, если будет найдена причина для того, чтобы энтропия начального состояния вселенной (сингулярности «большого взрыва») была сравнительно мала, а энтропия конечного состояния — велика. Согласно разд. 2.6, какое-то допущение о малой энтропии при «большом взрыве» действительно *необходимо*; иначе говоря, тот факт, что вселенная расширяется от некоторой сингулярности, сам по себе никак не достаточен. И наконец, согласно разд. 2.7, нам нужно наложить некоторые условия на начальные сингулярности, которые исключили бы сингулярности, находящиеся в центрах белых дыр. С другой стороны, рассуждения разд. 2.1 и 2.2 не сопровождались определенными выводами, и необходимо будет ненадолго вернуться к ним в конце.

Что же содержится в природе «большого взрыва» такого, из-за чего энтропия была бы «мала»? На первый взгляд должно показаться, что имеющиеся у нас знания о «большом взрыве» говорят о прямо противоположном. Материя (включая излучение) на ранних стадиях, по-видимому, полностью термализована (по крайней мере настолько, насколько это совместимо с расширением). Если бы это было не так, мы не получили бы правильного объяснения для обилия гелия и т. п. [77, 78]. Часто отмечалось, что численное значение «энтропии на один барион» (т. е. отношения числа фотонов к числу барионов) во вселенной «велико», а именно порядка 10^9.

Без учета вклада в энтропию от черных дыр эта величина, начиная с очень ранних стадий, в грубом приближении остается постоянной и составляет основную часть энтропии вселенной; но, несмотря на это, все «интересные» процессы, продолжающиеся в мире и столь важные для нашей жизни на Земле, обусловлены лишь «малым» приростом энтропии таких звезд, как наше Солнце. Ответ на мнимый парадокс, состоящий в том , что «большой взрыв» *рассматривается* нами как состояние с *высокой* энтропией, содержится в необычной природе гравитационной энтропии. Именно это я и собираюсь сейчас обсудить, а затем будет показано, какое отношение это имеет к структуре сингулярностей.

3.1. ГРАВИТАЦИОННАЯ ЭНТРОПИЯ

Многие авторы отмечали, что гравитация ведет себя в некоторой степени аномально в отношении энтропии [79]. Это справедливо как для ньютоновской теории, так и для общей теории относительности. (В действительности в ньютоновской теории положение даже хуже.) Так, во многих случаях, когда присутствует гравитация, система может вести себя, как если бы она имела отрицательную теплоемкость. Это явно имеет место в случае черной дыры, испускающей хокинговское излучение, поскольку чем больше она излучает, тем горячей она становится. Но даже в таких знакомых ситуациях, как движение спутника по орбите вокруг Земли, мы наблюдаем явления подобного рода. Так, диссипация (в виде трения в атмосфере) вызовет увеличение скорости спутника вместо ее уменьшения, т. е. приведет к росту кинетической энергии.

По существу, в этом проявляется универсально притягивающая природа гравитационного взаимодействия. По мере «релаксации» гравитирующей системы скорости возрастают, а источники скучиваются вместо равномерного распределения их по всему пространству с образованием более привычных высокоэнтропийных конфигураций. Другие виды сил в отношении притяжения имеют тенденцию к насыщению (как, например, в случае системы, связанной электромагнитно), но этого не происходит с тяготением. Только негравитационные силы могут воспрепятствовать дальнейшему коллапсу частей системы к их центрам по мере релаксации системы. Кинетическая энергия может лишь временно приостановить коллапс. В отсутствие существенных негравитационных сил, когда в игру вступают диссипативные явления, скучивание становится все более заметным, тогда как энтропия возрастает. Окончательно, при коллапсе с образованием черной дыры достигается максимальная энтропия, а это приводит нас снова к рассуждениям разд. 2.7.

Рассмотрим вселенную, которая расширяется от сингулярности «большого взрыва» и затем вновь коллапсирует (реколлапсирует) во всеохватывающую финальную сингулярность. Как утверждалось в

разд. 2.6, энтропия на поздних стадиях должна быть много больше энтропии ранних стадий. Каким образом этот рост энтропии проявляет себя? Какое различие между сингулярностями возникает из-за того, что энтропия финальной сингулярности велика, а энтропия «большого взрыва» относительно мала? Можно допустить, что первоначально *материя* имела большую энтропию, как это, видимо, и обстоит с реальной вселенной. Кинетической энергии «большого взрыва» (по крайней мере в среднем) вполне достаточно, чтобы преодолеть гравитационное притяжение, и вселенная расширяется. Но затем тяготение начинает неумолимо брать верх. Точный момент, когда это происходит в том или ином месте, зависит от степени уже имеющейся нерегулярности и, вероятно, от других разнообразных, но неизвестных факторов. Затем возникает скучивание, приводящее к появлению скоплений галактик, самих галактик, шарообразных скоплений, обычных звезд, планет, белых карликов, нейтронных звезд, черных дыр и т. д. Все знакомые нам сложные и интересные структуры обязаны своим существованием этому скучиванию, вследствие которого гравитационная потенциальная энергия начинает увеличиваться, а энтропия может начать расти выше *кажущегося* очень высоким значения, которое система имела первоначально. Следует ожидать, что это скучивание будет усиливаться; образуется все больше черных дыр, черные дыры поменьше будут поглощать вещество и сливаться друг с другом, образуя черные дыры большего размера. Этот процесс ускоряется в конечных стадиях реколлапса, когда средняя плотность опять становится очень большой и нужно ожидать, что конечное состояние будет крайне иррегулярным и зернистым.

Здесь мы встречаемся с определенной технической трудностью, состоящей в том, что понятие черной дыры обычно определяется только для асимптотически плоского (или открытого) пространства-времени. Эта трудность могла бы проявиться при рассмотрении конечных стадий коллапса, когда черные дыры начинают сливаться друг с другом и со всеохватывающей финальной сингулярностью реколлапса. Но в действительности нам нет необходимости знать локализацию горизонта черной дыры, а упомянутая трудность возникает лишь при точном определении этих горизонтов. Черная дыра, образовавшаяся в раннюю эпоху истории вселенной, имеет сингулярность, которую наблюдатели, встретившиеся с дырой, достигают при ранних значениях их собственного времени [57]; в сингулярности черных дыр, образовавшихся позднее, можно попасть в более поздние моменты. Исходя из строгой космической цензуры (см. разд. 3.2), нужно ожидать, что все эти сингулярности окончательно сольются с финальной сингулярностью реколлапса [57]. Я не требую, чтобы эти сингулярности можно было каким-то четким образом отличить друг от друга или от финальной сингулярности реколлапса. Важно лишь то, что гравитационное скучивание, которое характерно для состояния высокой гравитационной энтро-

пии, должно проявиться в виде крайне зернистой структуры конечной сингулярности (или сингулярностей).

Для вселенной, которая с момента своего «большого взрыва» расширяется бесконечно, картина вовсе не так сильно отличается от только что приведенной. По-прежнему следует ожидать локального скучивания и возникновения некоторого числа черных дыр (при условии, что начальная плотность не слишком мала и не слишком однородна для того, чтобы вообще было возможно образование галактик). Для областей внутри этих черных дыр ситуация несущественно отличается от ситуации в коллапсирующей вселенной (как это уже отмечалось в разд. 2.6), так что внутри каждой дыры, надо полагать, обнаружится весьма сложная сингулярность, соответствующая очень большой гравитационной энтропии. В областях, которые не являются внутренними для черных дыр, также будут определенные локализованные объекты, такие, как бесформенные глыбы, планеты, черные карлики, нейтронные звезды, которые соответствуют определенному предельному возрастанию уровня энтропии за счет гравитационного скучивания; однако прирост гравитационной энтропии в этих случаях будет сравнительно скромным, хотя и достаточным, как мы видим, для всего того, что требуется для жизни на Земле.

Я уже подчеркивал качественную связь между гравитационным скучиванием и увеличением энтропии вследствие роста гравитационной потенциальной энергии. В терминах пространственно-временной кривизны отсутствие скучивания соответствует (весьма приблизительно) нулевому значению конформной кривизны Вейля (поскольку отсутствие скучивания означает пространственную изотропию и, следовательно, отсутствие главных изотропных направлений) [45]. Когда имеет место скучивание, каждый сгусток окружен областью с отличной от нуля кривизной Вейля. С ростом сгустка гравитационное притяжение приводит к появлению новых областей с сильно возросшими значениями кривизны Вейля. Наконец, когда происходит гравитационный коллапс и образуется черная дыра, кривизна Вейля внутренней области становится еще больше и обращается в бесконечность на сингулярности.

Величина кривизны Вейля расходится как обратный куб расстояния от центра, по крайней мере в картине, описывающей сферически симметричный коллапс. Однако достаточно оснований и для уверенности в том, что и при коллапсе общего вида кривизна Вейля у сингулярности должна стремиться к бесконечности и доминировать над кривизной Риччи (в большинстве областей вблизи сингулярности).

Прямым свидетельством этого могут служить результаты анализа, проведенного Белинским, Халатниковым и Лифшицем [80]. Кроме того, к такому же выводу можно прийти на основе следующих чисто качественных рассуждений. Известно, что в точных фридмановских моделях доминирует тензор Риччи, а тензор Вейля пов-

сюду равен нулю. В этом случае если мировую линию частицы вещества продолжать до сингулярности, то она изотропно сближается с соседними такими же мировыми линиями, так что происходит одновременное схождение в трех взаимно перпендикулярных и ортогональных к мировой линии направлениях. В случае же сферически-симметричного коллапса к черной дыре, если мы следим за симметричным падением некоторого количества вещества на центральную сингулярность, мы, наоборот, увидим, что схождение к данной мировой линии обычно происходит лишь в *двух* взаимно перпендикулярных направлениях, ортогональных к этой мировой линии (а в третьем имеет место расхождение). Именно так обстоит дело в космологической модели Кантовского — Сакса [81], содержащей так называемую «сигарообразную» сингулярность [7]. Если r — обычная шварцшильдова координата, объем вблизи сингулярности спадает как $r^{3/2}$ и, следовательно, плотность — как $\sim r^{-3/2}$. Таким образом, типичная компонента Φ тензора Риччи ведет себя как $\Phi \sim r^{-3/2}$. В то же время типичная компонента Ψ тензора Вейля, вообще говоря, ведет себя как $\Psi \sim r^{-3}$. Отсюда видно, что в этих случаях тензор Вейля вблизи сингулярности доминирует. В сингулярностях типа «блина», где имеет место схождение только в *одном* направлении, ортогональном к мировой линии вещества, также следует ожидать, что тензор Вейля будет доминировать: в этом случае $\Phi \sim r^{-1}$ и $\Psi \sim r^{-2}$.

Таким образом, ситуация фридмановского типа с одновременным схождением всей материи и сразу со всех направлений представляется структурой весьма специального вида. Если схождение в одном направлении несколько меньше, чем в двух других, то, по-видимому, на очень близких расстояниях от сингулярности более вероятна сигарообразная конфигурация, в то время как блиноподобная появляется, видимо, когда схождение имеет место преимущественно только в одном направлении. Кроме того, при структуре общего вида, по всей вероятности, значительную роль играют осцилляции [80]. Осциллирующая кривизна Вейля с частотой ν и комплексной амплитудой Ψ дает дополнительный *эффективный* вклад «гравитационной энергии» в тензор Риччи порядка $|\Psi|^2\nu^{-2}$ [61]. Если ν становится очень большой, так что до достижения сингулярности происходит большое число осцилляций, то $\nu^2 \gg \Phi^{-1}$ [49], где Φ — типичная компонента тензора Риччи. Таким образом, если по мере приближения к сингулярности «энергетическое содержание» Ψ будет сравнимо с Φ (что, вообще говоря, представляется разумным), то будем иметь $|\Psi|^2\nu^{-2} \sim \Phi$ и, следовательно, $|\Psi| \gg \Phi$. Правда, эти рассуждения весьма приблизительны, но похоже, что они согласуются с более детальным анализом [80], из которого следует, что при рассмотрении поведения общего вида в первом приближении вкладом от материи можно пренебречь и трактовать решение как вакуумное, т. е. что вейлевская часть кривизны доминирует над кривизной Риччи.

Из всего этого следует, что высокоэнтропийная сингулярность связана с очень большой кривизной Вейля в отличие от заполненной пылью вселенной Фридмана и любых других моделей класса Робертсона — Уокера. Однако к моменту написания этой статьи еще не получено какой-либо четкой (скажем, интегральной) формулы, которую можно было бы рассматривать как математическое выражение этой предполагаемой связи между кривизной Вейля и гравитационной энтропией. Некоторые пути к этой формуле (если она вообще существует) могут подсказать, во-первых, формула Бекенштейна — Хокинга для энтропии черной дыры и, во-вторых, выражения для оператора числа частиц для свободного безмассового линейного квантованного поля спина 2, поскольку оценка «числа гравитонов» в гравитационном поле могла бы быть принята за меру его энтропии [1]). Таким образом, эта энтропия определяла бы число *квантовых* состояний, вносящих вклад в данную классическую геометрию.

В связи с вопросом об энтропии в гравитационном поле следует отметить еще один, последний пункт. Еще давно Толмен [84] указал, что в модельной вселенной, которая заполнена материей, находившейся на ранних стадиях в тепловом равновесии, может сложиться ситуация, когда эта материя в ходе расширения вселенной выйдет из равновесия (конкретным примером вещества, ведущего себя подобным образом, служит двухатомный газ, который способен к диссоциации на составляющие элементы и к рекомбинации). Тогда, если такая модель изображает расширяющуюся и реколлапсирующую вселенную, состояние материи в ходе реколлапса должно отличаться от соответствующего состояния в стадии расширения; соответствие мы устанавливаем по равенству значений радиуса вселенной R (сопутствующего радиуса). В самом деле, материя в ходе реколлапса должна получить некоторую энергию от глобальной геометрии вселенной, причем возникшее в результате этого различие в геометрии проявляется в том, что при данном R значение $\dot{R}^2$ больше при реколлапсе, чем при расширении. Поэтому энтропия системы как *целого* возрастает со временем несмотря на то, что *сама* материя находилась в начальной стадии расширения в тепловом равновесии. Фактически здесь имеет место вклад в энтропию от R (и $\dot{R}$), который должен рассматриваться в данной модели как динамическая переменная. (Этот вклад возникает за счет *объемной вязкости* [85].)

На происходящее в этой модели можно смотреть по существу как на преобразование потенциальной энергии глобальной структуры вселенной (гравитационная потенциальная энергия) в локаль-

[1]) Эта точка зрения, по-видимому, не согласуется с точкой зрения Гиббонса и Хокинга [83], которые, вероятно, считают, что гравитационная энтропия в отсутствие черных дыр равна нулю. Но оценка «числа фотонов» в классическом электромагнитном поле служит мерой его энтропии [83] (без черных дыр). С гравитацией дело обстоит, надо полагать, подобным же образом.

ную энергию вещества, хотя имеются известные трудности в точном определении энергии в моделях такого рода. Однако здесь эти трудности не должны нас особенно затрагивать, поскольку в действительности существенна энтропия, а не энергия, энтропия же значительно больше связана с вероятностями и структуризацией, чем с каким-либо частным определением энергии. В примере, приведенном Толменом, нет состояния с максимальной энтропией ни в рамках одной конкретной модели, ни для всех моделей этого типа вместе взятых. Выбирая значение R для максимального расширения достаточно большим (при фиксированном материальном содержимом), можно сделать полную энтропию сколь угодно большой. Толмен рассматривал возможность последовательных циклов «осциллирующей» вселенной с последовательно растущими максимальными значениями R. Нам, однако, теперь трудно придерживаться такого взгляда на вселенную ввиду теорем о сингулярностях [7, 8], утверждающих, что вселенная не может кончить свое эффективное «схлопывание» на некотором минимальном радиусе без нарушения известных нам законов физики [1]).

С моей точки зрения, ситуация, обрисованная Толменом, может рассматриваться как один из аспектов вопроса о том, каким образом структура вселенной как целого вносит вклад в энтропию. Здесь мы, очевидно, касаемся несколько иного аспекта этого вопроса, чем при рассмотрении гравитационного скучивания, поскольку тензор Вейля в моделях Толмена всюду равен нулю. Ясно, что в этом тоже необходимо детально разобраться, если мы хотим полностью понять роль гравитационной энтропии. Тем не менее достаточно ясно, что прирост энтропии в ситуациях типа рассмотренных Толменом относительно незначителен [78] по сравнению с тем приростом, который может быть обнаружен и *в самом деле* обнаруживается при гравитационном скучивании (ср. с разд. 3.3).

Ключевым вопросом, который должен интересовать нас в наибольшей степени, является структура сингулярностей. Эти сингулярности по меньшей мере определяют граничные условия для разных циклов толменовской «осциллирующей» вселенной. Более того, как мы очень скоро увидим, наличие иррегулярностей не должно изменить всеохватывающего характера этих сингулярностей в случае расширяющейся и реколлапсирующей вселенной, если выполняется строгая космическая цензура.

3.2. КОСМИЧЕСКАЯ ЦЕНЗУРА[2])

Хотя допущение о том, что «голые» сингулярности недопустимы, ни в коей мере не существенны для точки зрения, которую я соби-

---

[1]) Я рассматриваю квантовую гравитацию как «неизвестную нам» независимо от того, насколько она помогает решению проблемы сингулярностей!

[2]) Некоторые части этого раздела значительно более сложны в технических деталях, чем остальные, но их можно пропустить без потери основной нити рассуждений.

раюсь выдвинуть, все же такого рода предположение о «космической цензуре» весьма упрощает рассуждения, и они становятся яснее. Я уже заявил, что по большинству вопросов буду стремиться следовать общепринятому взгляду, так что не было бы отклонением от этой линии, если бы я просто принял некоторую подходящую формулировку предположения о космической цензуре (по крайней мере как предмет для обсуждения), равняясь при этом на то, что представляется мне взглядом большинства. Но все же на следующих страницах я приведу некоторое независимое оправдание этому взгляду.

Однако прежде чем переходить к деталям, необходимо сделать одно замечание. В хокинговском процессе испарения черной дыры при (предполагаемом) финальном исчезновении дыры на мгновение возникает «голая» сингулярность. Это обычно не рассматривается как нарушение космической цензуры, поскольку процесс Хокинга — квантовомеханический процесс, в то время как обычно считается, что космическая цензура относится лишь к классической общей теории относительности. (По словам Хокинга [68], космическая цензура здесь скорее «преодолевается», чем нарушается.) Тем не менее наличие реальной «голой» сингулярности такого рода в геометрии мира внесло бы определенные изменения в наши рассуждения. Но похоже, что эти изменения не имели бы какого-либо серьезного отношения к рассматриваемой проблеме. Все черные дыры, верить в реальное существование которых во Вселенной можно на достаточно серьезных основаниях, должны иметь массы порядка солнечной или больше, а мы знаем, что времена жизни таких черных дыр больше 10^{53} времен Хаббла, и поэтому можем спокойно игнорировать их финальные «голые» сингулярности. Кроме того, с той точки зрения, которую я выдвинул в разд. 3.3, существование мини-дыр крайне маловероятно, а это единственные черные дыры (скажем, массой 10^{20} г и меньше), финальные «голые» сингулярности которых появятся достаточно рано для того, чтобы иметь хотя бы отдаленную связь с нашим рассмотрением. Все же поскольку в любом случае такие дыры были бы не больше размеров атома, их можно было бы «сгладить» и не рассматривать в качестве существенной составной части классической геометрии.

С учетом всего этого обсуждение космической цензуры целиком в рамках классической общей теории относительности, по-видимому, будет вполне соответствовать нашим целям. Итак, в чем состоит утверждение о космической цензуре, которого мы должны придерживаться? Обычная формулировка представляет собой утверждение типа:

В системе, которая эволюционирует, согласно общей теории относительности, от несингулярных начальных данных общего вида на подходящей гиперповерхности Коши, при разумных уравне-

ниях состояния не появится какая-либо пространственно-временна́я сингулярность, которая была бы видна с бесконечности.

Нечто в этом роде, запрещающее «голые» сингулярности, оказывается необходимым для того, чтобы можно было получить обычные общие результаты по черным дырам (например, принцип возрастания площадей, слияние двух дыр с необходимым появлением третьей, общая макроустойчивость черных дыр, вплоть до утверждения о физическом появлении именно черных дыр, а не чего-то еще худшего при коллапсе общего вида [51]). Приведенная формулировка лишена ясности в нескольких пунктах, и потребовались бы значительные уточнения, чтобы извлечь из нее что-либо доступное ясному математическому доказательству или опровержению.

Вероятно, нам все же мало помогут какие-либо уточнения тех или иных условий, если не будет более глубокого представления о том, в чем же именно состоит дело. Например, мне кажется совершенно неразумным предполагать, что физика в сравнительно локальной области пространства-времени должна в самом деле «заботиться» о том, «вырвется» ли луч света, испущенный из сингулярности, на «бесконечность» или нет. Можно рассудить иначе: некий наблюдатель (времениподобная мировая линия) мог бы пересечь этот луч света и увидеть данную сингулярность как «обнаженную», хотя он в действительности не находился бы на бесконечности (ни один реальный наблюдатель никак не может там находиться). Наблюдатель может находиться вблизи сингулярности и, возможно, даже быть захвачен ею, т. е. находиться внутри обычной черной дыры (см. рис. 5, *а*). Наблюдатель, видящий эту «локально обнаженную» сингулярность, в не меньшей мере, чем наблюдатель на бесконечности, столкнулся бы с непредсказуемостью, которой сопровождается присутствие обнаженной сингулярности и которая многим так претит.

Мне кажется, что довольно несущественно, может ли сам наблюдатель «вырваться» на бесконечность. Классическая общая теория относительности является масштабно-независимой теорией, и потому, если локально обнаженные сингулярности наблюдаются при очень малых расстояниях, то они в принципе должны наблюдаться и при очень больших расстояниях, когда «захваченный» наблюдатель мог бы иметь дни и даже годы для размышлений о последствиях неопределенностей, возникших в результате его наблюдения такого рода сингулярности (ср. с аналогичным рассуждением об астронавте внутри огромной черной дыры в разд. 2.6). В самом деле, для обитателей реколлапсирующих замкнутых вселенных (каковыми, возможно, являемся и мы сами) нет никакой «бесконечности», так что для них в принципе нет вопроса, «захвачены» они или «не захвачены»,— локально можно говорить лишь о степени «захваченности».

Таким образом, если космическая цензура является принципом Природы, то, по-видимому, она должна быть сформулирована так, чтобы такого рода *локально* обнаженные сингулярности были устранены [57, 58, 86]. Эта точка зрения находит определенное подтверждение прежде всего в стандартной картине сферически-симметричного коллапса внутри черной дыры, изображенной на рис. 5, *а*. В самом деле, наблюдатель, падающий внутрь этой черной дыры, вообще не может увидеть сингулярность, пока не столкнется с ней. Возможно, это станет более очевидным, если использовать для описания ситуации стандартную конформную диаграмму (с изотропными конусами, проведенными под углом 45°), как на рис. 8, поскольку тогда пространственноподобный характер сингулярности выступает явственно, и видно, что в описанном выше смысле локально обнаженной сингулярности *нет*.

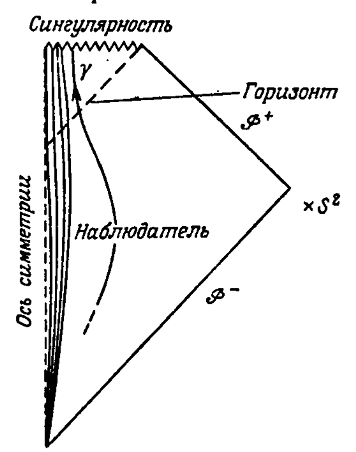


Рис. 8. Конформная диаграмма сферически-симметричного коллапса (в асимптотически плоском пространстве-времени), иллюстрирующая пространственноподобный характер сингулярности.

Кроме того, есть основания считать, что возмущения общего вида, снимающие сферическую симметрию, не изменят пространственноподобного характера сингулярности (тот факт, что она будет продолжать существовать и в возмущенном случае следует из теорем сингулярности [6—8]). Ситуация, однако, несколько усложняется вследствие того, что сингулярность Шварцшильда — Крускала (рис. 8) в действительности неустойчива. Добавление ничтожного по величине момента количества движения в черную дыру приведет к решению Керра (при $a \ll m$), что фактически полностью изменит структуру сингулярности, и она перестанет быть пространственноподобной. Только при дальнейшем возмущении общего вида можно ожидать восстановления пространственноподобной структуры сингулярности.

Несколько легче исследовать поведение сингулярности, если мы рассмотрим добавление заряда, а не момента количества движения; тогда мы получим решение Райсснера — Нордстрема вместо решения Керра. Соответствующая конформная диаграмма изображена на рис. 9, *а*, откуда видно, что теперь сингулярность в самом деле *является* локально обнаженной в приведенном выше смысле, поскольку наблюдатель, мировая линия которого γ, может увидеть сингулярность. При дальнейшем возмущении общего характера надо ожидать положения, более напоминающего изображенное на рис. 9, *б*, где сингулярность вновь пространственноподобна

(или изотропна). Причина здесь в том, что в ситуации, изображенной на рис. 9, *а*, в пространстве-времени имеется изотропная гиперповерхность *H*, которая является *горизонтом Коши* пространственноподобной гиперповерхности (без края) Σ, простирающейся до пространственной бесконечности. (Терминологию и обозначения [1]) см. в работе [49].) Наблюдатель γ, который пересечет

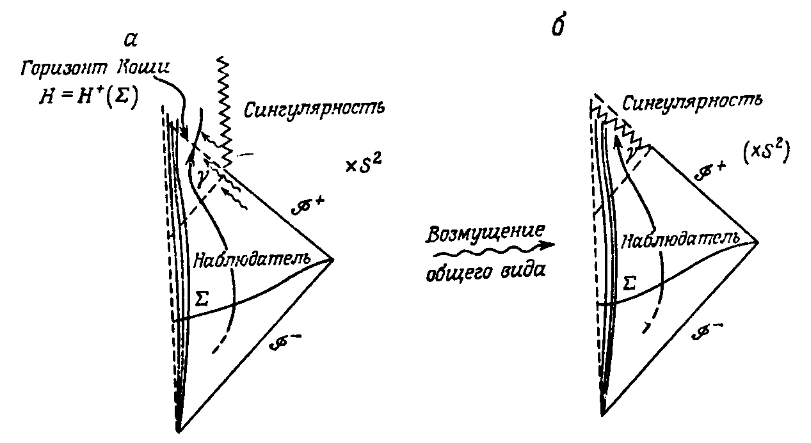


Рис. 9. Конформные диаграммы, иллюстрирующие коллапс в черную дыру с малым зарядом.
а — сферически-симметричный; *б* — при возмущении общего вида.

$H = H^{+}(\Sigma)$, увидит всю будущую историю мира снаружи как бы сжатую в мгновенную вспышку. Если данные на Σ будут каким-то образом немного изменены вблизи бесконечности, то это приведет к решительному изменению геометрии в окрестности *H*, обусловленному тем, что сигналы из бесконечно удаленных областей Σ претерпят вблизи *H* фиолетовое смещение бесконечной величины. Действительно, анализ возмущений слабыми (пробными) полями [87, 88] ясно показывает, что эти возмущения приводят к расходимости вдоль *H*. При полном нелинейном взаимодействии можно ожидать, что на месте *H* будет локализована сингулярность кривизны. Более того, вполне *могло бы* оказаться, что нелинейные эффекты большой кривизны в действительности приведут к пространственноподобной, а не к изотропной сингулярности, поскольку эти эффекты большой кривизны могут порождать и усиливать друг друга все более и более, пока не возникнет сингулярность (ср. рис. 9, *б*).

Чтобы придать точность понятию пространственноподобной (или изотропной) сингулярности, стоит напомнить о понятии *идеальной точки* [89] пространства-времени $\mathscr{M}$, определенной в тер-

[1]) Читатель может обратиться также к книге Хокинга и Эллиса [7], где терминология и обозначения в основном те же.— *Прим. перев.*

минах множеств ГНП или ГНБ (граничные неразложимые множества прошлого или будущего) для этого $\mathscr{M}$. Идеальные точки можно представлять себе как некоторые «добавочные» точки, присоединенные к многообразию $\mathscr{M}$ как «сингулярности» или как «точки на бесконечности», которые служат для того, чтобы времениподобные кривые в $\mathscr{M}$, не имеющие будущих концевых точек в $\mathscr{M}$, снабдить будущими идеальными концевыми точками (с помощью ГНП), а кривые, не имеющие прошлых концевых точек в $\mathscr{M}$,— прошлыми идеальными концевыми точками (с помощью ГНБ).

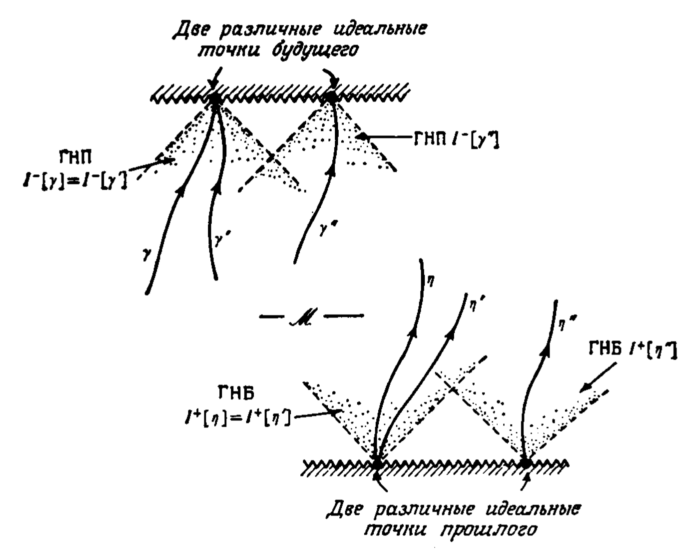


Рис. 10. Множества ГНБ и ГНП, определяющие идеальные точки пространственно-временного многообразия $\mathscr{M}$.

Предположим для простоты, что $\mathscr{M}$ — сильно причинное многообразие. Пусть γ и γ'— две времениподобные кривые в $\mathscr{M}$ без будущих концевых точек. Тогда γ и γ' обладают одной и той же будущей идеальной концевой точкой, если и только если они имеют одинаковое прошлое, что в стандартных обозначениях записывается как $I^-[\gamma]=I^-[\gamma']$. Множества ГНП многообразия $\mathscr{M}$ являются по существу множествами вида $I^-[\gamma]$ при времениподобной γ без будущей концевой точки, и их можно «отождествить» с будущими идеальными точками. Подобным же образом η и η' времениподобные кривые без прошлой концевой точки имеют одинаковые прошлые идеальные концевые точки, если только будущее каждой из них совпадает с будущим другой: $I^+[\eta]=I^+[\eta']$ и эти множества являются множествами ГНБ многообразия $\mathscr{M}$ (рис. 10). В каждом из этих случаев говорят, что данное ГНП или ГНБ *порождено* рассматриваемой времениподобной кривой. Простой критерий [57,

58], который может быть применен для того, чтобы отличить ГНП, изображающие точки на бесконечности, от ГНП, изображающих сингулярные точки, состоит в определении ГНП как ∞-ГНП, если

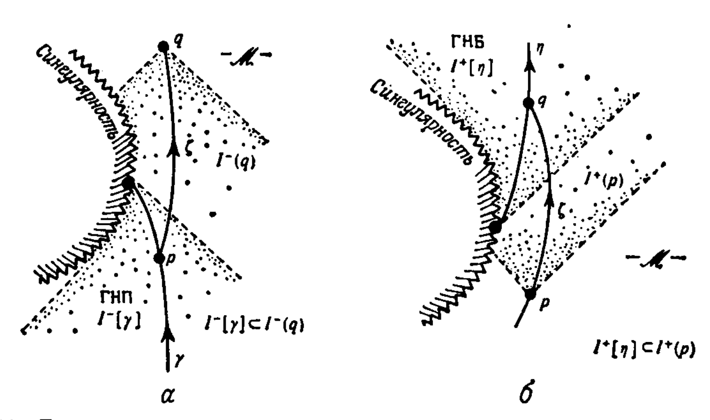


Рис. 11. Локально обнаженная сингулярность, лежащая в будущем точки p и в прошлом точки q.
а — определение через ГНП; *б* — определение через ГНБ.

оно порождено времениподобной кривой бесконечной будущей собственной длины, и как *сингулярное* ГНП, если среди порождающих кривых нет кривой бесконечной будущей собственной длины; ∞-ГНБ и *сингулярное* ГНБ определяются аналогично. (Сингулярными в определенном смысле можно было бы называть также некоторые из ∞-ГНП и ∞-ГНБ, но я не буду вникать здесь в такие детали.)

Далее, локально *обнаженную* сингулярность можно определить либо как сингулярную ГНП, содержащуюся в прошлом $I^-(q)$ некоторой точки $q \in \mathcal{M}$, либо как сингулярную ГНБ, содержащуюся в будущем $I^+(p)$ некоторой $p \in \mathcal{M}$ (рис. 11). В результате такого определения в каждом из этих двух случаев имеется времениподобная кривая ζ (мировая линия наблюдателя) из точки p в точку q, такая, что сингулярность лежит в будущем p и в прошлом q. (Возьмите p в ГНП в первом случае и q в ГНБ — во втором.) Тут важно не только то, что некий наблюдатель (а именно ζ) может увидеть сингулярность (а именно из q), но также и то, что к некоторому более раннему моменту его существования сингулярности уже должны возникнуть. Таким образом, «обычный» всеобъемлющий «большой взрыв» *не* квалифицируется как локально обнаженная сингулярность, поскольку никаких наблюдателей не существовало до того, как он произошел.

Можно сказать, что $\mathcal{M}$ согласуется с определенного рода космической цензурой, если таких локально обнаженных сингулярностей (т. е. типа изображенных на рис. 11) нет [57]. Однако полезно пойти несколько дальше и исключить также «обнаженные

точки на бесконечности» путем отказа от условия, чтобы ГНП или ГНБ, имеющиеся на рис. 11, были непременно *сингулярными* ГНП или ГНБ. Конечно, можно было бы утверждать, что такие обнаженные точки на бесконечности вносят столько же неопределенности в будущее поведение модели вселенной, как и обнаженные сингулярности. Однако на самом деле возникновение таких бесконечно удаленных обнаженных точек, по-видимому, неправдоподобно, если только они не являются в некотором смысле *сингулярными* точками (хотя они по-прежнему будут определяться множествами ∞-ГНП или ∞-ГНБ). Причина заключается в том, что при гладкой конформной бесконечности $\mathcal{I}$ обнаженные точки на бесконечности существуют только при условии, что $\mathcal{I}$ (по крайней мере местами) времениподобна. Но при разумной плотности материи $\mathcal{I}$ может быть времениподобной, только если $\lambda<0$, где λ — космологическая постоянная [90]. Однако обычные модели с $\lambda<0$ (типа фридмановской) расширяются от всеохватывающей сингулярности и к такой же сингулярности снова сжимаются [91]. Следовательно, в них нет ни ∞-ГНП, ни ∞-ГНБ, а потому сингулярных обнаженных точек на бесконечности даже в таких случаях не следует ожидать.

Независимо от того, соответствуют ГНП или ГНБ сингулярным точкам или точкам на бесконечности, из ситуации, изображенной на рис. 11, видно, что идеальные точки принадлежат границе $\mathcal{M}$, которая в определенном смысле *времениподобна*. Это объясняется следующим образом. Если мы продолжим ζ неограниченно в будущее (рис. 11, *а*), так чтобы она стала времениподобной кривой ζ' без концевой точки в будущем, то получим

$$I^-[\gamma]\subset I^-(q),\quad q\in I^-[\zeta'],$$

а это есть условие [89] того, что ГНП $I^-[\gamma]$ лежит в хронологическом (т. е. времениподобном) прошлом ГНП $I^-[\zeta']$. (Более слабое условие $I^-[\gamma]\subset I^-[\zeta']$ сводится к тому, что $I^-[\gamma]$ лежит в *причинном* прошлом множества $I^-[\zeta']$ [89].) Аналогично рис. 11, *б* дает

$$I^+[\eta]\subset I^+[p],\qquad p\in I^+[\zeta''],$$

что при ζ'' без концевой точки в прошлом характеризует ГНП $I^+[\eta]$ как лежащее в хронологическом будущем множества ГНБ $I^+[\zeta'']$. Таким образом, исключение конфигурации, изображенной на рис. 11, *а*, эквивалентно утверждению, что будущие идеальные точки составляют *ахрональную* [49] (т. е., не вполне точно говоря, пространственноподобную или изотропную) будущую границу многообразия $\mathcal{M}$, в то время как исключение конфигурации, изображенной на рис. 11, *б*, эквивалентно утверждению, что прошлые идеальные точки составляют *ахрональную* прошлую границу $\mathcal{M}$.

Более того, исключение только *одной* из этих конфигураций во всем $\mathcal{M}$ эквивалентно исключению и другой, поскольку каждое

из условий оказывается эквивалентным [57] симметричному по времени условию *глобальной гиперболичности* $\mathcal{M}$ [7, 49].

Доказательство этого утверждения совсем несложно, и поскольку оно не приводится в литературе, стоит дать здесь его набросок. Во-первых, глобально-гиперболическое пространство-время — это пространство-время с тем свойством, что для любых двух его точек p, q пространство причинных кривых от p до q компактно. (Причинной называется кривая, не обязательно гладкая, которая везде локально проходит внутри светового конуса будущего или по нему. Таким образом, причинные кривые времениподобны или всюду являются локально C^0-*пределами* времениподобных кривых.) В предположении сильной причинности глобальная гиперболичность эквивалентна утверждению, что каждое $I^+(q) \cap I^-(p)$ обладает компактным замыканием [7, 49]. Я отмечу здесь также, что для причинной кривой ζ множество $I^-[\zeta]$ есть ГНП, если и только если ζ не имеет будущей концевой точки [89]; аналогичное утверждение имеет место для ГНБ.

Теперь предположим, что $\mathcal{M}$ содержит точку q и времениподобную кривую γ без будущей концевой точки, такую, что $I^-[\gamma] \subset \subset I^-[q]$. Отсюда следует, что $\mathcal{M}$ не может быть глобально-гиперболическим многообразием, ибо если p — фиксированная точка на γ и r_1, r_2, r_3, ... — последовательность точек, распространяющаяся бесконечно далеко по γ, мы получим последовательность ζ_1, ζ_2, ζ_3, ... причинных кривых от p до q, в которой ζ_i состоит из сегмента кривой γ от p до r_i и некоторой времениподобной кривой от r_i до q [которая существует, поскольку $r_i \in I^-[\gamma]$ и $I^-[\gamma] \subset I^-(q)$]. Если бы ζ_1, ζ_2, ... имели предельную причинную кривую ζ от p до q, то выполнялось бы $I^-[\zeta'] = I^-[\gamma]$, где (как это легко видеть) $\zeta' = \zeta \cap I^-[\gamma]$ ввиду того, что ζ' не может быть кривой без будущей концевой точки; ее можно продлить по ζ к будущей концевой точке q. Такой вывод противоречил бы последнему утверждению предыдущего абзаца, откуда следует, что ζ не существует и, следовательно, $\mathcal{M}$ не глобально-гиперболично.

Обратное утверждение доказывается технически более сложно. Я пользуюсь обозначениями и нумерацией теорем из работы [49]. Допустим, $\mathcal{M}$ не является глобально-гиперболическим. Тогда существуют точки p и q, для которых $I^+(p) \cap I^-(q)$ не обладает компактным замыканием. Отсюда (предложения 3.9, 5.20 в работе [49]) $p \notin \operatorname{int} D^-(\partial I^-(q))$, если учесть, что $p \in I^-(q)$. Следовательно (предложение 5.5*h*), $p' \notin D^-(\partial I^-(q))$, где $p' \in I^-(p)$, так что имеется времениподобная кривая γ без будущей концевой точки, выходящая из p' и не пересекающая $\partial I^-(q)$. Но $p' \in I^-(q)$, откуда следует, что $\gamma \subset I^-(q)$ и ГНП $I^-[\gamma]$ полностью лежит внутри $I^-(q)$, что и требовалось доказать.

Мое предложение [57, 59] относительно *принципа строгой космической цензуры* состоит, следовательно, в том, что физически приемлемое классическое пространство-время $\mathcal{M}$ должно обладать

свойством, которое можно определить любым из следующих эквивалентных утверждений: ни одно ГНП не лежит целиком в прошлом какой-либо точки $\mathcal{M}$; ни одно ГНБ не лежит целиком в будущем какой-либо точки $\mathcal{M}$; все ГНП образуют ахрональное множество; все ГНБ образуют ахрональное множество; $\mathcal{M}$ глобально гиперболично; в $\mathcal{M}$ существует гиперповерхность Коши. (Эквивалентность двух последних утверждений — известный результат, принадлежащий Героку [93].) Правдоподобие всего этого зависит, естественно, от наших представлений о том, что́ такое «физически приемлемое» пространство-время. Действительно, глобальную гиперболичность многие склонны считать слишком сильным ограничением. Тем не менее я верю, что *можно* выдвинуть внушающие доверие аргументы в пользу строгой космической цензуры. Сейчас я их кратко намечу.

Представляется разумным для начала ограничиться только случаем вакуума (хотя при этом исключается случай «большого взрыва»). Основания для этого были отмечены в разд. 3.1, а именно, следует ожидать, что вблизи сингулярности «общего вида» тензор Вейля доминирует над тензором Риччи. Это, конечно, не совсем удовлетворительно из-за тех «кумулятивных» эффектов (фокусирование), которые обусловлены тензором Риччи, а тензором Вейля могут создаваться лишь косвенным путем, через нелинейности. Тем не менее поведение вакуумных решений, по-видимому, должно служить хорошим первым приближением вблизи сингулярности общего вида, позволяющим избежать тех проблем, которые возникают, например, из-за явно несущественных «проникающих сквозь оболочку» «голых» сингулярностей [92] в идеализированной материи типа «пыли». В качестве второго приближения можно было бы рассмотреть, например, теорию Эйнштейна — Максвелла, в которой также нет подобных проблем. Что касается «большого взрыва», то он представляет собой особую ситуацию (что связано с его низкой энтропией), и критерий «общности» здесь неприменим [1]). Все же строгая космическая цензура и здесь, по-видимому, действует, но по другим причинам (ср. с разд. 3.3). В случае же сингулярностей коллапса высокоэнтропийное предположение об «общности» представляется физически разумным.

Последовательность ситуаций, изображенных на рис. 8, 9, *а*, *б*, по-видимому, должна служить одной из вероятных путеводных нитей к общей ситуации. На рис. 8 (продолженное решение Шварцшильда) условие глобальной гиперболичности выполняется, но явно случайно. Каждое сингулярное ГНП пересекает поверхность Коши Σ по компактной области. Данные в *одной* только этой области — это все, что необходимо для доказательства существования и описания характера сингулярности, которая этим ГНП опреде-

[1]) Это в сущности вполне естественно, ибо «большой взрыв» — нечто такое, что при обращении времени было бы *неустойчивой* сингулярностью (ср. с разд. 3.3).

ляется. Однако при небольшом возмущении этого решения, таком, чтобы оно стало решением Керра, и при его максимальном продолжении обычным образом (как на рис. 9, *а*) сингулярность исчезает в том смысле, что теперь не существует никакого сингулярного ГНП вблизи первоначального сингулярного ГНП, которое заменилось прошлым $I^-(x)$ несингулярной внутренней точки x. Таким образом, первоначальная сингулярность явно обязана своим существованием какой-то специфике (например, точной сферической симметрии) исходных начальных данных. Глобальная гиперболичность в слегка возмущенном решении нарушена, но вследствие этого нарушения теперь имеется горизонт Коши $H=H^+(\Sigma)$. Прошлое $I^-(y)$ точки $y \in H$ пересекает теперь Σ по области с *некомпактным* замыканием (простирающейся до бесконечности). Мы можем принять, что структура пространства-времени (например, кривизна) в окрестности y определяется некоторым интегралом от начальных данных по этой области. Если на эти данные накладывается возмущение общего вида, мы можем получить *расходимость* (за счет некомпактности и возникающего вследствие этого бесконечного фиолетового смещения), так что несингулярная точка y превращается в итоге в сингулярную идеальную точку с (предположительно) бесконечной кривизной.

Допустим, что вместо асимптотически плоского случая, который мы до сих пор изучали, рассматривается такая начальная пространственноподобная гиперповерхность Σ, которая компактна. По-прежнему может оказаться, что, подобно ситуации, изображенной на рис. 9, *а*, определенные наборы данных на Σ приведут к максимально расширенному вакуумному пространству-времени, в котором нарушается условие глобальной гиперболичности (например, к пространству Тауба — НУТ) и возникает горизонт Коши $H=H^+(\Sigma)$. Возьмем, как прежде, $y \in H$ и рассмотрим $I^-(y) \cap \Sigma$. Это множество должно иметь теперь компактное замыкание (поскольку Σ компактна), но, по-видимому, оно в определенном смысле обязано быть *эффективно-некомпактным* вследствие бесконечного «наматывания» Σ на $I^-(y)$; по крайней мере именно об этом говорит изучение моделей типа Бианки IX [80, 94]. Эта эффективная некомпактность должна была бы, по-видимому, привести к ситуации, похожей на только что рассмотренную, в которой при возмущении общего вида H превращается в сингулярность кривизны и выполняется строгая космическая цензура, поскольку интегралы, определяющие возмущенную кривизну на H, учитывают одни и те же данные на Σ снова и снова бесконечное число раз.

Возникает соблазн заключить из этого, что сингулярности типа изображенных на рис. 8, которые ведут свое происхождение от данных на эффективно-компактной области, представляют собой особый случай «меры нуль» и что, когда определенные возмущения приводят к возникновению горизонтов Коши, точки на этих горизонтах обязаны зависеть от эффективно-некомпактных носителей данных Коши, причем фиолетовые смещения будут бесконечными,

вследствие чего горизонты должны быть неустойчивыми (как на рис. 9, *а*). В этом смысле предположение о строгой космической цензуре выглядит весьма правдоподобным [1]); однако приведенное здесь рассуждение весьма далеко от того, чтобы быть доказательством.

Итак, мы стоим теперь перед картиной глобально-гиперболической вселенной, которая начинается с некоторого ахронального множества $\partial\check{\mathcal{M}}$ начальных идеальных точек («большой взрыв»), затем топологически остается неизменной (следствие глобальной гиперболичности [93]), несмотря на присутствие черных дыр, до тех пор, пока не будет достигнуто ахрональное множество $\partial\hat{\mathcal{M}}$ финальных идеальных точек. (Строгая космическая цензура означает фактически, что $\partial\check{\mathcal{M}}$ и $\partial\hat{\mathcal{M}}$ должны рассматриваться как абсолютно непересекающиеся множества.) Начальные идеальные точки все считаются сингулярными точками, но финальные идеальные точки могут быть как бесконечно удаленными, так и сингулярными точками. Можно думать, что бесконечно удаленные точки возникают лишь в случае модели вечно расширяющейся вселенной, причем в этом случае следовало бы ожидать также и наличия сингулярных конечных идеальных точек в черных дырах. Вместе с тем допустимо предположение (хотя, на мой взгляд, это довольно маловероятно по причинам, упомянутым выше), что во вселенной, которая как целое реколлапсирует, будут иметься какие-то ограниченные области, которые «вырвутся» на бесконечность, в (несингулярное) ∞-ГНП.

На рис. 12 изображена модель бесконечно расширяющейся вселенной, и мы видим, как многообразие $\mathcal{M}$ может остаться топологически неизменным в том смысле, что $\mathcal{M} \cong \mathbb{R}\times\Sigma$, причем каждая копия $\mathbb{R}$ есть времениподобная кривая, а каждая копия Σ — пространственноподобная гиперповерхность Коши [93], несмотря на возможное присутствие некоторого числа черных дыр [57]. Ситуация с моделью реколлапсирующей вселенной аналогична. Множества $\partial\check{\mathcal{M}}$ и $\partial\hat{\mathcal{M}}$, однако, могут иметь, но могут и не иметь ту же топологию, что и Σ. (Например, в статической вселенной Эйнштейна каждое из множеств $\partial\check{\mathcal{M}}$ и $\partial\hat{\mathcal{M}}$ есть точка, тогда как Σ есть S^3.) Согласно точке зрения, изложенной в разд. 3.3, $\partial\check{\mathcal{M}}$ для «большого взрыва» должно в действительности обладать топологией Σ. Но совсем не очевидно, что такой же должна быть и топология $\partial\hat{\mathcal{M}}$. Высказанная первоначально Мизнером надежда на то, что в «перемешанных» вакуумных моделях общего вида типа Бианки IX не будет горизонтов частиц [94], может быть перефразирована

[1]) По ряду различных соображений с точки зрения *математики* весьма желательно иметь возможность ограничиться рассмотрением глобально-гиперболических пространственно-временных многообразий. При этом, например, исчезает большинство технических трудностей, касающихся топологии и отождествлений в структурах ГНП и ГНБ [89].

в том смысле, что $\partial\check{\mathcal{M}}$ должно представлять собой единственную точку [1]). При обращении времени можно было бы ожидать, что и $\partial\hat{\mathcal{M}}$ будет единственной точкой. Но дальнейший анализ [95—98] показал, что поведение такого рода представляет собой крайне маловероятную возможность. Тем не менее можно думать, что при

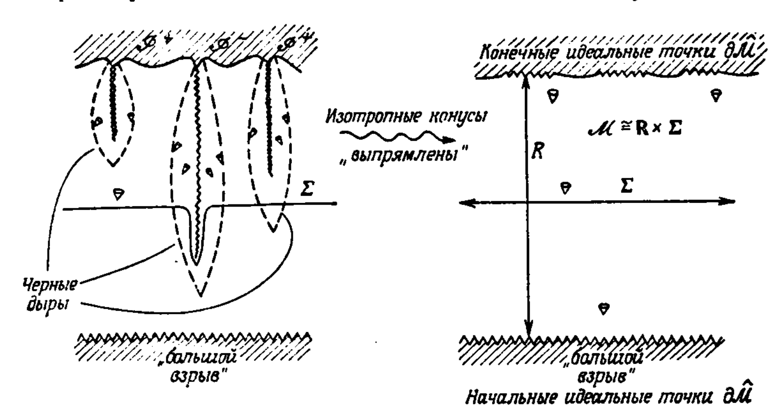


Рис. 12. Модель вселенной, в которой действует строгая космическая цензура. Иллюстрируется случай бесконечного расширения. (Различные области $\mathcal{J}^+$ в действительности все связаны между собой.)

определенных условиях и в определенном смысле число измерений $\partial\hat{\mathcal{M}}$ будет меньше трех (как в случае двумерных $\partial\hat{\mathcal{M}}$ для сингулярностей типа «блина» и для пространства Тауба [7]).

Если $\partial\check{\mathcal{M}}$ для «большого взрыва» действительно является гладкой пространственноподобной гиперповерхностью с топологией гиперповерхности Σ (ср. с разд. 3.3), то она представляет собой весьма подходящую начальную гиперповерхность Коши. Но безотносительно к гладкости или трехмерности множеств $\partial\check{\mathcal{M}}$ и $\partial\hat{\mathcal{M}}$ каждое из них может рассматриваться как область задания предельных данных Коши для $\mathcal{M}$, а именно $\partial\check{\mathcal{M}}$ — область задания предельных начальных данных и $\partial\hat{\mathcal{M}}$ — предельных конечных данных. Множества $\partial\check{\mathcal{M}}$ и $\partial\hat{\mathcal{M}}$ являются всеохватывающими в том смысле, что каждое из них пересекается любой причинной кривой в $\mathcal{M}$ без концевой точки; кроме того, они ахрональны. Конечно, пока нет ясности в том, какой вид могли бы иметь эти данные Коши, но в принципе определенные перспективы здесь имеются.

Этот раздел мы завершим изучением возможности начального возникновения и финального исчезновения заряженной частицы (ср. также с разд. 2.7). Если начальное рождение в окрестности $\partial\check{\mathcal{M}}$ можно считать результатом более или менее понятного нам

[1]) То же самое и при том более явно требуется для «детерминистического» пространства-времени Сакса — Будича [99].

процесса [62, 63, 66], при котором кривизна сама создает частицы, то следовало бы ожидать появления не одной заряженной частицы, а *пары* противоположно заряженных частиц. Однако могло статься, что в рождение частиц при «большом взрыве» примешивался и совершенно неизвестный нам процесс, приводящий к появлению заряженных частиц поодиночке [67]. В этом случае в окрестности $\partial\breve{\mathcal{M}}$ должно также появиться кулоновское поле этой частицы, ведущее, как было отмечено в разд. 2.4, к присутствию некоторого эффективного приходящего излучения без источников [45]. Конечно, такую картину (по крайней мере на первый взгляд) можно было бы рассматривать как обращение во времени исчезновения частиц в окрестности $\partial\hat{\mathcal{M}}$. Представим себе одиночную частицу, которую поглотила большая сферически-симметричная незаряженная черная дыра. Согласно геометрии, представленной на рис. 8, эта частица должна аннигилировать в одиночестве в сингулярной части множества $\partial\hat{\mathcal{M}}$. Можно ли допустить, что эта единственная частица в такой степени изменит геометрию $\partial\hat{\mathcal{M}}$, что ее исчезновение задержится до тех пор, пока черная дыра не захватит частицу противоположного знака, не направит ее к первому заряду для аннигиляции, и только после этого энергия обеих частиц будет поглощена сингулярностью? Думается, в это трудно поверить. Все же структура сингулярностей общего вида представляется столь хитроумно и сложно устроенной, что даже такую возможность нельзя упускать из виду. Однако вопрос, который я пытаюсь здесь поднять, состоит в следующем: могут ли все-таки быть симметричными по времени законы, управляющие физическим поведением в окрестности пространственно-временнóй сингулярности? На мой взгляд, те, кто считает ответ положительным, могут столкнуться с серьезными дополнительными проблемами принципиального характера.

3.3. МОЯ ГИПОТЕЗА И ПРЕДСТАВЛЕНИЕ О МИРЕ

Почти все рассуждения, которые здесь выдвигались, сфокусированы, как мне кажется, на одном вопросе: что представляет собой та особая структура, присущая «большому взрыву», которая отличает его от обращенной во времени сингулярности коллапса общего вида, и почему эта структура именно такая? Здесь сразу следует напомнить точку зрения *хаотической космологии* [94], которая была в моде в течение ряда лет. Согласно этой точке зрения, «большой взрыв» сначала представлял собой не слишком однородную сингулярность, но пришел к однородному состоянию вследствие того, что диссипативные явления (например, вязкость нейтрино, столкновения адронов или рождение частиц [100—102]) сгладили все его иррегулярности. Одним из предвкушаемых следствий этой диссипации должно было стать наблюдаемое «большое» значение

энтропии на барион ($\sim 10^9$). В качестве другого следствия должна была получиться наблюдающаяся на нынешней стадии изотропия. А для того чтобы не накладывать на «большой взрыв» явно произвольных ограничений, в качестве исходного было выдвинуто предположение, что хаос начальной сингулярности был в некотором подходящем смысле *максимальным*. На этом пути был предложен некий предполагаемый канонический вид начального состояния, из которого можно было вывести подробные следствия и сравнить их с данными наблюдений.

Должен сказать, однако, что программу хаотической космологии (по крайней мере в ее «чистой» форме максимального начального хаоса) я считаю основанной на недопонимании. Ведь утверждать, что начальный хаос максимален,— это, по-видимому, то же самое, что утверждать, будто начальная *энтропия* максимальна. А если это так, то при симметричных по времени физических законах не было бы никакой стрелы времени, а следовательно, и никакой диссипации. (Апелляция к расширению вселенной здесь ничего не дает по причинам, которые мы пространно обсуждали в разд. 2.6!)

Мы могли бы, однако, заняться несколько более умеренным вариантом хаотической космологии, в котором начальный геометрический хаос не был бы максимальным, а был подходящим образом ограничен, причем ограничения были бы устроены так, чтобы можно было получить из ранней диссипации корректно как наблюдаемое значение *энтропии* на барион $\sim 10^9$, так и ту плотность иррегулярностей, которая могла бы объяснить образование галактик как следствие неоднородности начальной геометрии [103, 105]. На мой взгляд, однако, и этот вариант хаотической космологии неудовлетворителен. Существенно неверным пунктом является, по-видимому, то, что энтропия порядка 10^9 на барион считается *большой* в гравитационном контексте. Рассмотрим замкнутую реколлапсирующую вселенную, в которой содержится, скажем, 10^{80} барионов. Когда образуется черная дыра массой $M_\odot$, ее энтропия на барион по формуле Бекенштейна — Хокинга достигает порядка 10^{18}, а галактика массой 10^{10} $M_\odot$ с черной дырой массой 10^6 $M_\odot$ в центре имеет энтропию на барион порядка 10^{20}. Когда вся масса галактики сколлапсирует в дыру, эта цифра возрастет до 10^{28}. По мере развития коллапса вселенной эти черные дыры объединяются в новые, все бóльшие по размерам дыры, и при этом финальная энтропия на барион достигает чудовищной величины — порядка 10^{40}. Эта энтропия «оседает» на иррегулярностях геометрии финальной сингулярности. Обращая теперь направление времени, мы видим, какая удручающе огромная энтропия получилась бы, если бы Творец предпочел воспользоваться хаотической начальной геометрией. Считающаяся «большой» цифра $\sim 10^9$ на самом деле мелочь по сравнению с этим. Если для объяснения цифры 10^9 придерживаться «умеренно» хаотической космологии, то мы долж-

ны спросить, почему на самом деле используется лишь ничтожно малая часть (не больше 10^{-31}) всего допустимого хаоса?! (В действительности эта часть будет всего лишь порядка 10^{-32}, если для энтропии на барион в замкнутой вселенной принять величину порядка 10^8.)

Похоже, что при столь большом несоответствии нам не приходится рассчитывать на гравитационное объяснение цифры 10^9 [1]). Более обнадеживающей областью поисков ее объяснения могла бы стать физика элементарных частиц. В разд. 4 я вернусь к этому вопросу. Подобным же образом *чисто* гравитационное объяснение иррегулярностей, необходимых для образования галактик, представляется не слишком перспективным. И опять (хотя и со значительно меньшей уверенностью) я рассчитываю на физику элементарных частиц — на самых ранних стадиях расширения.

Таким образом, я исхожу из того, что в начальной *геометрии* хаос должен полностью отсутствовать [57—60]. Нам нужно, во всяком случае, какое-то низкоэнтропийное ограничение на начальное состояние. Однако *материя* (включая излучение) на ранних стадиях находилась в тепловом равновесии (или была близка к нему). Так что «малость» начальной энтропии была результатом не какого-то специального распределения материи, а скорее результатом некой весьма специальной начальной геометрии пространства-времени. Рассуждения в разд. 2.6 и 3.1 указывают, в частности, что это ограничение на раннюю геометрию должно быть чем-то вроде утверждения: *кривизна Вейля C_{abcd} обращается в нуль в окрестности любой начальной сингулярности* [2]) [58—60].

Эта гипотеза пока несколько туманна и открывает возможности для ряда отличных друг от друга интерпретаций. Мы могли бы потребовать, например, чтобы кривизна Вейля просто стремилась к нулю по мере приближения к начальной сингулярности, или задать некоторую скорость стремления, или же, возможно, потребовать лишь, чтобы она оставалась ограниченной (или даже чтобы тензор Риччи просто доминировал вблизи сингулярности так, чтобы тензор кривизны становился в пределе *пропорциональным* тензору с исчезающей вейлевской частью). Я пока не исследовал вопрос о том, какие различия могут возникнуть при разных вариантах этой гипотезы. Попытаемся сначала обратиться к простейшему варианту, когда $C_{abcd} \to 0$ (скажем, в любой параллельно распространяющейся системе отсчета) по мере приближения к начальной сингулярности. Отмечу вкратце, какие приблизительно следствия могли бы здесь иметь место, хотя полную картину еще только предстоит установить.

---

[1]) Это подтверждается недавним подробным анализом Барроу и Матцнера [104], которые пришли по существу к тем же выводам.

[2]) Иными словами (разд. 3.2), в окрестности $\partial_0\mathscr{M}$. Заметим, что это относится и к конечному сингулярному ГНБ хокинговского взрыва черной дыры, в котором кривизна Вейля действительно обращается в нуль.

Позволю себе сначала разъяснить роль этой гипотезы в том «выборе», который Творец делает, отбирая *конкретную* вселенную из имеющегося у него, по-видимому, бесконечного набора вселенных, согласующихся с данными физическими законами. Представим себе некоторое «обширное» многообразие $\mathcal{U}_0$ (здесь я пользуюсь словом «многообразие» в несколько вольном смысле), изображающее всевозможные различные начальные данные для вселенных, совместимые с физическими законами. Чтобы выбрать одну из вселенных, Творец просто накалывает «булавку» где-то в $\mathcal{U}_0$. Но точка зрения нашей статьи состоит в том, что при выборе граничных условий мы не должны отдавать предпочтение начальным данным перед конечными. Поэтому с равным успехом на выбор вселенной можно было бы смотреть как на накалывание Творцом булавки в многообразии $\mathcal{U}_\infty$, представляющем собой всевозможные комплекты конечных данных, совместимых с физическими законами. Кроме того, можно было бы воспользоваться любым промежуточным $\mathcal{U}_t$, состоящим из всевозможных данных «в момент времени t» [1]). Все они эквивалентны, а уравнения движения (о которых в рамках нашего обсуждения предполагается попросту, что они классически детерминированного типа) устанавливают канонический изоморфизм между $\mathcal{U}_0$, $\mathcal{U}_\infty$ и каждым из $\mathcal{U}_t$. Таким образом, мы можем рассматривать *одно* изоморфное абстрактное многообразие $\mathcal{U}$, которое представляет любое (или каждое) из этих многообразий и является множеством всех возможных историй вселенных, согласующихся с физическими законами.

Мы должны набраться мужества и согласиться с тем, что булавка Творца наколота в $\mathcal{U}$ «случайным образом» (ибо если бы ее положение было ограничено каким-либо контролируемым способом, то это ограничение составило бы еще один «закон физики»). Но понятие «случайно» требует того, чтобы в $\mathcal{U}$ была введена некоторая подходящая *мера.* Является ли эта «случайность» одним и тем же понятием в приложении как к начальным, так и к конечным состояниям? Поставим этот вопрос иначе: совпадает ли мера фазового пространства, естественным образом введенная на $\mathcal{U}_0$, с мерой, заданной каноническим изоморфизмом на $\mathcal{U}_\infty$ (а также на каждом $\mathcal{U}_t$)? Теорема Лиувилля говорит нам, что *совпадает* при условии, что мы придерживаемся общепринятой гамильтоновой физики, а я не собираюсь быть «оригиналом» в этом отношении. Более того, я собираюсь игнорировать любые трудности [2]), какие

[1]) Понятие «время» используется здесь весьма вольно. В правой части рис. 12 (где не обязательно нужно предполагать, что расширение бесконечно) t могло бы быть некоторым параметром, пробегающим значения от 0 до ∞ и измеряющим «высоту» вверх по рисунку.

[2]) Я должен принести извинения, особенно специалистам, за мою сырую и легковесную трактовку столь деликатного предмета, каким я занялся. Оправданием мне может быть лишь уверенность в том, что в вопросах, которые в данный момент рассматриваются, общерелятивистские и термодинамические усложнения, на мой взгляд, не являются ключевыми пунктами.

могут возникнуть из возможной бесконечномерности или бесконечности полной меры многообразия $\mathcal{U}$, не говоря уж обо всех весьма серьезных «калибровочных» проблемах (и т. п.), которые возникли бы при настоящем общерелятивистском рассмотрении.

Каким образом должна рассматриваться с этой точки зрения энтропия? Стандартная процедура состоит в наделении каждого $\mathcal{U}_t$ крупнозернистой структурой путем деления $\mathcal{U}_t$ на «зерна», в которых различные элементы любого одного «зерна» соответствовали бы состояниям, макроскопически неразличимым между собой в момент t. Если булавка Творца протыкает $\mathcal{U}_t$ в точке, принадлежащей «зерну» фазового пространства объемом V (в единицах, в которых $\hbar=1$), то (больцмановская) энтропия в момент t равна $k \ln V$ (ср. с разд. 2.3). Эта энтропия может флуктуировать [1]) или постепенно меняться со временем, так как крупнозернистая структура для разных $\mathcal{U}_t$ не отображается одна на другую при канонических изоморфизмах. Малая энтропия в один момент (точка в малом «зерне») может соответствовать большой энтропии в другой момент (точка в большом «зерне»), где «специализированность» состояния переходит теперь в макроскопически неразличимые корреляции.

В рассматриваемом контексте, однако, такое описание энтропии все же не удовлетворительно. Точки $\mathcal{U}$ соответствуют только тем историям вселенных, которые совместны с физическими законами во *все* моменты времени. Можно думать, что в момент t макроскопически нельзя установить, удовлетворяются ли эти законы во все остальные моменты времени. Вот тут-то я и выдвигаю гипотезу о том, что в действительности имеются (локальные) физические законы, которые становятся важными только вблизи сингулярностей пространства-времени, причем они асимметричны по времени и обеспечивают обращение в нуль кривизны Вейля в любой начальной сингулярной точке (т. е. в точках множества $\partial\breve{\mathcal{M}}$). Действие этих законов проявляется в том, что каждое многообразие $\mathcal{U}_t$ оказывается много меньше, чем можно было бы ожидать без них. Допускаются только такие движения и конфигурации в момент t, которые согласуются с ограничением $C_{abcd}=0$ при $t=0$. Однако поскольку при этом на каждое $\mathcal{U}_t$ накладываются ограничения, которые не являются макроскопически различимыми, для вычисления энтропии в момент t неразумно использовать просто объемы фазового пространства в пределах $\mathcal{U}_t$. Вместо этого мы должны рассматривать расширенные объемы в некотором более широком многообразии $\mathcal{W}_t$, определенном так же, как $\mathcal{U}_t$, но для которого не требуется выполнения этих (начальных) ограничений.

---

[1]) Если предпочесть описание мира в терминах *ансамблей*, то можно в определенной мере избежать этих флуктуаций, считая, что Творец пользуется *тупой* булавкой! Пока диаметр острия булавки больше, чем структурное зерно, флуктуации оказываются сглаженными.

Уравнения движения снова задают изоморфизмы между различными многообразиями $\mathscr{W}_t$ в различные моменты t (по крайней мере локально и постольку, поскольку добавочные, ограничивающие законы остаются физически несущественными), и снова существует соответствующее абстрактное многообразие $\mathscr{W}$, представляющее всю совокупность не связанных ограничениями историй вселенных. Вложение $\mathscr{U}$ в $\mathscr{W}$ имеет вполне конкретную связь с крупнозернистой структурой при $t=0$, поскольку обращение в нуль кривизны Вейля *является* макроскопически различимым свойством. Поэтому никакое деление $\mathscr{U}$ при $t=0$ не расширяется за пределы $\mathscr{U}$, в $\mathscr{W}$. Но по мере роста t соответствующие «зерна», образующие крупнозернистую структуру $\mathscr{U}$, расширяются все более и более в $\mathscr{W}$ и соответственно каждое «зерно» приобретает все больший объем [1]). С этой точки зрения мы воспринимаем «специфичность» современного состояния вселенной (возникшую из-за того, что вселенная стартовала с $C_{abcd}=0$) как приобретающую со временем все более и более характер «точных корреляций» и все более теряющую «низкоэнтропийный» характер.

Это и приводит к соответствию с принципом возрастания энтропии в нашей Вселенной. При таком подходе мы не делаем никаких статистических низкоэнтропийных предположений относительно «большого взрыва», а вместо этого накладываем строгое *локальное условие* $C_{abcd}=0$. В остальном допускается полная случайность, т. е. Творец накалывает булавку в многообразии $\mathscr{U}$ случайным образом. При этом предположении о случайности мы можем приписать «причину» отсутствия начальных корреляций в движении частиц в начальном состоянии (т. е. *закона условной независимости* [40]) тому факту, что новое локальное ограничение не наложено ни в какой другой момент, кроме момента $t=0$ (например, никакого условия не накладывается на финальную сингулярность $t=\infty$). Соответственно «причиной» растущих корреляций при росте t и роста энтропии является начальное ограничение $C_{abcd}=0$. Таким способом проблема стрелы времени может быть выведена из области статистической физики и возвращена в ту область, где занимаются установлением точных (локальных?) физических законов. Этот вопрос будет кратко обсужден в разд. 4.

Здесь следует упомянуть об интенсивно обсуждавшемся *антропологическом* принципе [106], к которому часто обращаются в связи с поднятыми мной вопросами. Этот принцип в сущности должен был бы означать, что булавка Творца втыкается в $\mathscr{U}$ не совсем случайно, а с некоторым весовым множителем, отдающим предпочтение вселенным, содержащим (много?) *разумных наблю-*

---

[1]) Любопытно, что если бы не происходило этих расширений в $\mathscr{W}$, то объемы наибольших «зерен» со временем *уменьшались* бы из-за избытка различных геометрий, которые при этом нарождались бы. Это, по-видимому, соответствует тому факту, что вселенная может стать более «интересной», несмотря на рост энтропии.

дателей. (Более того, булавка могла бы быть воткнута также и в другие многообразия $\mathcal{V}^{o}$, $\mathcal{W}$, . . ., соответствующие всевозможным другим альтернативным наборам числовых значений физических постоянных и всевозможным альтернативным законам физики. Я отказываюсь от обсуждения этого обширного вопроса как «лежащего за пределами предмета данной статьи»!) Такого рода антропологический принцип время от времени привлекали, когда пытались объяснить неравновесие энтропии в наблюдаемой Вселенной, приписывая больший вес той гигантской флуктуации, которая могла бы понадобиться для создания условий, необходимых для жизни [107]. Трудность здесь заключается в том, что несравненно «дешевле» (в единицах отрицательной энтропии) просто породить несколько разумных существ в некотором тщательно организованном столкновении частиц, чем создать за счет флуктуации такое неравновесие энтропии, какое, как нам, землянам, известно, имеется во *всей* по видимости безграничной Вселенной, по крайней мере если судить о ней по области, достижимой для наиболее мощных телескопов!

Этим я не хочу сказать, будто уверен, что предположение о «случайном» характере движения острия булавки Творца всегда будет наилучшим объяснением нынешнего состояния мира. Однако с добавлением предположения о равенстве нулю начального значения C_{abcd} оно, по-видимому, работает удивительно хорошо и хотя бы на некоторое время освобождает нас от забот о глубоком смысле всего происходящего с точки зрения физики.

Итак, какие же следуют выводы из той картины мира, которую я предлагаю? Коль скоро первоначально кривизна Вейля равна нулю, а вещество (и излучение) находится в тепловом равновесии, то в начальном состоянии у нас будет нечто весьма близкое к пространственной изотропии и однородности. Таким образом, подход и результаты Фридмана, Робертсона и Уокера вполне оправдываются (на начальной стадии) и ведут к поразительному согласию с рядом замечательных наблюдений: с однородностью фонового излучения черного тела температурой 2,7 К (с точностью $\sim 10^{-3}$) [108], с отсутствием доступного измерению вращения Вселенной по отношению к инерциальным системам отсчета ($<10^{-16}$ с$^{-1}$) [77], с однородностью объемного распределения галактических скоплений на больших расстояниях. Сама сингулярность «большого взрыва» должна быть при этом близка к робертсоно-уокеровскому типу [1]). *Определенные* флуктуации распределения вещества с этой точки зрения допускаются, и в действительности они даже *должны* быть, поскольку начальные ограничения на R_{ab} являются статистическими в отличие от ограничений на C_{abcd}. Однако первоначальное равенство нулю C_{abcd} налагает серьезные ограничения на

---

[1]) Независимым подтверждением этому служит точность расчетов по образованию гелия [77, 78].

такие начальные флуктуации плотности и скоростей. Необходимо лучше разобраться в физике частиц, прежде чем можно будет провести вычисления этих флуктуаций во всех деталях [109].

Белые дыры исключаются во все моменты времени, поскольку их сингулярности являются сингулярностями «начального» типа (т. е. точками $\partial\check{\mathcal{M}}$), которые даже приблизительно не удовлетворяют условию $C_{abcd}=0$. Черные дыры, конечно, допускаются при условии, что они образуются обычным способом вследствие гравитационного коллапса массивного тела или тел. Но мини-дыр, по-видимому, *не* должно быть, поскольку они требуют начального состояния с хаотической геометрией. Несуществование таких первичных черных дыр согласуется с современными наблюдениями [110].

Наконец, главным и наиболее сильным следствием такой картины мира становится наблюдаемое экстраординарное поведение энтропии нашего мира, которое настолько пронизывает весь наш повседневный опыт, что мы склонны считать его само собой разумеющимся.

4. АСИММЕТРИЧНАЯ ФИЗИКА?

Некоторые читатели могут почувствовать себя разочарованными всем этим. Вместо того чтобы изыскать какой-то тонкий ход к тому, чтобы во вселенной, основанной на симметричных по времени законах, могла тем не менее наблюдаться явная асимметрия по времени, я попросту утверждаю, что некоторые из этих законов в действительности несимметричны по времени, и, хуже того, утверждаю, что эти асимметричные законы пока еще неизвестны! Но на самом деле все не так плохо, как может показаться. В частности, из всего этого следует, что стоит поискать такие асимметрии и в других областях физики. Где именно? Так или иначе, некоторая связь с гравитацией при этом должна быть, поскольку именно тензор Вейля описывает гравитационные степени свободы. Классическая общая теория относительности симметрична по времени, но можно поставить вопрос: сохранится ли эта симметрия, когда в конце концов теория гравитации будет должным образом «сшита» с квантовой механикой? В самом деле, если считать, что виртуальные черные дыры при планковской длине (10^{-33} см) играют существенную роль [73], то аргументы разд. 2.7 приводят к тому, что вакуум может быть асимметричным по времени в рамках сугубо квантового описания гравитации.

Однако это не слишком доступный и обнадеживающий путь, но есть иная заслуживающая исследования возможность связи между квантовой механикой и гравитацией, которая, вероятно, имеет больше отношения к вопросу об асимметрии по времени — это вопрос о *квантовомеханических наблюдениях*, отложенный нами в разд. 2.2. Здесь следует подчеркнуть определенные ассоциации

с подходом Бекенштейна — Хокинга к энтропии. В конечном счете квантовомеханическое наблюдение рассматривается как «свершившееся», лишь когда происходит нечто «необратимое». Но слово «необратимость» относится здесь к тому факту, что произошло существенное увеличение энтропии. А энтропия, как мы уже знаем, зависит, по-видимому, от довольно субъективного понятия крупнозернистой структуры. Для того чтобы квантовомеханическое наблюдение «действительно» произошло и привело к реальному изменению в состоянии мира, по-видимому, требуется, чтобы увеличение энтропии было *объективным*. Вспомним теперь, что в формуле Бекенштейна — Хокинга в качестве меры энтропии взята попросту конкретная характеристика пространственно-временной геометрии — площадь поверхности черной дыры. Означает ли это, что и геометрия дыры тоже субъективна с вытекающим отсюда выводом, что *вся* пространственно-временна́я геометрия (а значит, и вся физика) должна быть в какой-то мере субъективной? Или же энтропия в случае черной дыры должна стать объективным понятием? Если последнее верно, то не может ли энтропия стать объективной также и в менее экстремальных гравитационных ситуациях (ср. с разд. 3.1)? Более того, если такой переход от субъективного характера энтропии к объективному может произойти только в (квантовой) теории гравитации, то именно в ней окончательно теряет силу линейная суперпозиция цепей фон Неймана [111]!

В этом случае возможно, что новые законы, вероятно, необходимые для такого обобщения квантовой механики, при котором измерения станут неотъемлемой частью теории, приведут к определенного рода *квантовой теории гравитации* (см. разд. 2.2), причем под этим я имею в виду теорию, из которой как соответствующие предельные случаи следуют квантовая механика и общая теория относительности. Во всяком случае, я бы сказал, что приводимые мною аргументы (особенно те из них, которые содержатся в разд. 2.7 и 3.3 и наиболее непосредственно связаны с формулой Бекенштейна — Хокинга) ставят вопрос о некоторой новой теории, *асимметричной* по времени. Соответственно, какая бы нелинейная физика [1]) ни заменила в конечном счете нынешнюю теорию с мгновенно редуцируемой волновой функцией, она вполне может оказаться существенно асимметричной по времени.

На нынешнем уровне знаний такие рассуждения в высшей степени спекулятивны. Все же мы знаем, что некий физический

---

[1]) Недавно Бялыницкий-Бируля и Мысельский [112] предложили интересную модификацию уравнения Шредингера, добавив к нему член $(b \ln |\psi|^2)\psi$. Хотя этот член не является асимметричным по времени, он все же имеет связь с обсуждаемым нами вопросом: значение постоянной b (температура) порядка 10^{-8} K, что совпадает с температурой по Бекенштейну — Хокингу для черной дыры массой 10 $M_\odot$. Как раз таковы наименьшие из черных дыр, в существование которых можно верить на достаточных физических основаниях.

закон, асимметричный по времени, реально *существует*! За более привычными симметричными по времени силами Природы где-то скрыта одна (а возможно, и не одна) асимметричная сила, ничтожное действие которой почти полностью замаскировано остальными и остается незамеченным во всех процессах, кроме одного: хитроумного распада K^0-мезона. Я не хочу сказать, что здесь обязательно замешана квантовая гравитация. Однако, очевидно, что в числе нерушимых правил Природы *нет* требования непременной симметрии по времени!

Более того, в относительной величине T-нарушающей (или CP-нарушающей) компоненты в K^0-распаде (около 10^{-9}), возможно, имеется свой скрытый смысл. Согласно разд. 3.3, физика элементарных частиц должна как-то объяснить наблюдаемое значение энтропии на один барион порядка 10^9. В связи с этим высказывалась мысль [62, 67, 109], что в ранней Вселенной, возможно, шел процесс рождения барионов и антибарионов, в котором превышение барионов над антибарионами составило примерно $1+10^{-9}:1$. Тогда при последующей аннигиляции получилось бы не только (видимо) требующееся число барионов, но и наблюдаемая энтропия на барион (т. е. число фотонов, приходящихся на один барион). Если можно считать, что исходным состоянием, с которого начал развиваться этот первоначальный процесс рождения, было нечто с квантовыми числами вакуума, то в этом процессе должен нарушаться закон сохранения барионов, а также CP- и C-инвариантности [1]) (чтобы число барионов и антибарионов было неодинаково). Значительное нарушение C-инвариантности, как уже давно известно, является свойством слабых взаимодействий [20], тогда как CP-инвариантность нарушается в K^0-распаде в столь малой степени, какая, по-видимому, и требуется. Более того, основываясь на хокинговском процессе испарения черной дыры, следует ожидать, что несохранение барионов [76] и в настоящее время *где-то* имеет место [56, 67].

Однако для объяснения закона, утверждающего, что первоначально $C_{abcd}=0$, нам нужно нечто большее, а именно нарушение T-, PT-, CT- и CPT-инвариантностей по отдельности. Если бы, например, CPT-инвариантность не нарушалась, мы могли бы взять незапрещенный коллапс к сингулярности, для которой $C_{abcd}\neq 0$, и по CPT-симметрии получить запрещенную начальную сингулярность. PT- и CP-инвариантности максимально нарушены в слабых взаимодействиях, а T-инвариантность — еле заметна в K^0-распаде, но CPT-нарушения пока еще не были обнаружены. Конечно, имеется CPT-теорема [113], которая служит определенной теоретической базой для универсальной CPT-инвариантности, но не следует

---

[1]) Ввиду асимметрии по времени в расширяющейся вселенной можно также предполагать нарушение CPT- или CT-инвариантности вместо соответственно CP- или C-инвариантности.

забывать, что важнейшей предпосылкой этой теоремы является пуанкаре-инвариантность, тогда как мы обращаемся к ситуациям (сингулярности, квантовая гравитация), в которых это условие явно нарушено. Я мог бы поспорить, что какие-то *CPT*-нарушающие эффекты в конце концов в экспериментальной физике будут замечены. Но пока еще рано рассчитывать на это, и ни в коей мере не удивительно, что такие эффекты мы до сих пор не видим.

Здесь я не вдавался в какие-либо количественные детали. Однако большинство явлений, которых мы касаемся, носит столь всеобщий и явный характер, что о многом можно судить и без деталей. Из этих явлений более всего выделяется статистическая асимметрия вселенной. По-моему, совершенно непостижимо, чтобы такая асимметрия не имела осязаемой причины. Объяснение с помощью антропологического принципа представляется крайне неубедительным (ср. с разд. 3.3). То же можно сказать и об объяснении «нарушением симметрии» [114], согласно которому наиболее вероятные состояния вселенной могут не разделять тех симметрий, которыми обладают управляющие вселенной законы. (Трудно понять, как наша безбрежная Вселенная могла «впасть» именно в то или другое из этих состояний, не имея даже понятия, в каком направлении времени стартовать!) Насколько я сам могу судить, остается *единственное* («очевидное») объяснение: *не все* точные законы физики симметричны по времени!

Тогда возникает вопрос: для чего Природе понадобилось запрятать эту асимметрию по времени столь тщательно? Поскольку мы пока не знаем тех принципов, которыми руководствуется Природа при отборе физических законов, мы не можем ответить и на этот вопрос. Все же нам, по-видимому, не следует слишком уж удивляться тому, что фундаментальная асимметрия глубоко скрыта за внешностью видимой симметрии. Обладает же фауна нашей Земли, за редкими исключениями, внешней симметрией между правым и левым. Можно ли было догадаться, что в ядре каждой воспроизводящей клетки заложена спираль, структура которой определяет рост и развитие этих великолепных симметричных созданий, но при том каждая такая спираль все-таки только правая?

БЛАГОДАРНОСТИ

Я благодарен Деннису Шаме и особенно Амелии Речел-Кон за критическое прочтение рукописи и за то, что они обратили мое внимание на ряд работ других авторов. Приношу благодарность также Стивену Хокингу за несколько бесед, которые мне многое прояснили.

ЛИТЕРАТУРА

1. *Einstein A.*, *Rosen N.*, Phys. Rev. ser. 2, 48, 73 (1935).
2. *Einstein A.*, Ann. Math., 40, 922 (1939).
3. *Einstein A.*, Berl. Ber. 235 (1931).
4. *Einstein A.*, Ann. Math., ser. 2, 46, 578 (1945).
5. *Einstein A.*, *Straus E.*, Ann. Math., ser. 2, 47, 731 (1946).
6. *Penrose R.*, Phys. Rev. Lett., 14, 57 (1965).
7. *Hawking S. W.*, *Ellis G. F. R.*, The Large Scale Structure of Space-Time, Cambridge University Press, 1973. [Имеется перевод: *Хокинг С.*, *Эллис Дж.* Крупномасштабная структура пространства-времени.— М.: Мир, 1977.]
8. *Hawking S. W.*, *Penrose R.*, Proc. R. Soc. Lond., A314, 529 (1970).
9. *Davies P. C. W.*, The Physics of Time Asymmetry. Surrey University Press, 1974.
10. *Grunbaum A.*, Philosophical Problems of Space and Time. Knopf, New York, 1963.
11. *Reichenbach H.*, The Direction of Time. University of California Press, Berkeley, 1956.
12. *Einstein A.*, *Tolman R. C.*, *Podolsky B.* Phys. Rev., ser. 2, 37, 780 (1934).
13. *Watanabe S.*, Rev. Mod. Phys., 27, 179 (1955); Prog. Theor. Phys., Suppl. (Extra No.), 135 (1965).
14. *Aharanov Y.*, *Bergmann P. G.*, *Lebowitz J. L.*, Phys. Rev., B134, 1410 (1964).
15. *Penfield R. H.*, Am. J. Phys., 34, 422 (1966).
16. *D'Espagnat B.*, The Conceptual Foundations of Quantum Mechanics. Benjamin, Menlo Park, California, 1971.
17. *Christenson J. H. et al.*, Phys. Rev. Lett., 13, 138 (1964).
18. *Casella R. S.*, Phys. Rev. Lett., 21, 1128 (1968); 22, 554 (1969).
19. *Kabir P. K.*, The *CP* Puzzle. Academic Press, London, New York, 1968.
20. *Commins E. D.*, Weak Interactions. McGraw-Hill, New York, 1973.
21. *Wolfenstein L.*, Phys. Rev. Lett., 13, 562 (1964).
22. *Einstein A.*, *Podolsky B.*, *Rosen N.*, Phys. Rev., ser. 2, 47, 777 (1935).
23. *Einstein A.*, Ideas and Opinions. Souvenir Press, London, 1974.
24. *Schrödinger E.*, Naturwiss., 23, 807, 823, 844 (1935).
25. *Wigner E. P.*, In Symmetries and Reflections, MIT Press, 1970; and in The Scientist Speculates, ed. I. J. Good. Heineman, London, 1961.
26. *Everett H. III*, Rev. Mod. Phys., 29, 454 (1957).
27. *Wheeler J. A.*, Rev. Mod. Phys., 29, 463 (1957).
28. *DeWitt B. S.*, In: Foundations of Quantum Mechanics, ed. B. d'Espagnat. Academic Press, London, New York, 1971.
29. *Clarke C. J. S.*, Phil. Sci., 41, 317 (1974).
30. *Bohr N.*, Atomic Theory and the Description of Nature. Cambridge University. Press, 1934, 1961.
31. *Dirac P. A. M.*, In: The Physicist's Conception of Nature, ed. J. Mehra, p. 14. Reidel. Dordrecht, 1973.
32. *Komar A.*, Int. J. Theor. Phys., 2, 157 (1969).
33. *Karolyhazy F.*, Nuovo Cimento, A42, 390 (1966).
34. *Born M.*, Nature, 119, 354 (1927).
35. *Feynman R. P.*, *Leighton R. B.*, *Sands M.*, The Feynman Lectures on Physics, vol. III. Addison-Wesley, Reading, Massachusetts, 1963. [Имеется перевод; *Фейнман Р.*, *Лейтон Р.*, *Сэндс М.* Фейнмановские лекции по физике. Т. 3—4.— М.: Мир, 1976.]
36. *Ehrenfest P.*, *Ehrenfest T.*, The Conceptual Foundations of the Statistical Approach in Mechanics. Cornell University Press, 1959.
37. *Tolman R. C.*, The Principles of Statistical Mechanics, Oxford University Press, 1938.
38. *Penrose O.*, Foundations of Statistical Mechanics. Pergamon, Oxford, 1970.
39. *Hahn E. L.* Phys. Rev., 80, 580 (1950).
40. *Penrose O.*, *Percival I. C.*, Proc. Phys. Soc., 79, 509 (1962).
41. *Gold T.*, The Nature of Time, Cornell University Press, 1967.

42. *Hogarth J.*, Proc. R. Soc. Lond., **A267**, 365 (1962).
43. *Sciama D. W.*, Proc. R. Soc. Lond., **A273**, 484 (1963).
44. *Einstein A.*, *Ritz W.*, Physikalische Zeitschrift, 9 Jahrgang, no. 25, 903; 10 Jahrgang, no. 6, 185; 10 Jahrgang, no. 9. 322.
45. *Penrose R.*, In: Relativity, Groups and Topology, eds. C. M. DeWitt, B. S. DeWitt, Gordon & Breach; New York, 1964; см. также [41]. [Имеется перевод в сб. статей: «Гравитация и топология».— М.: Мир, 1966.]
46. *Geroch R. P.*, In: Battelle Rencontres, eds. C. M. DeWitt and J. A. Wheeler. Benjamin, New York, 1968.
47. *Cf. Lord Dunsany*, The King That Was Not, in Time and the Gods.
48. *Gold T.*, Am. J. Phys., **30**, 403 (1962).
49. *Penrose R.*, Techniques of Differential Topology in Relativity. S. I. A. M. Philadelphia, 1972.
50. *Bell J. S.*, *Perring J. K.*, Phys. Rev. Lett., **13**, 348 (1964).
51. *Penrose R.*, Rivista Nuovo Cimento, Num. Spec. I, **1**, 252 (1969).
52. *Новиков И. Д.*, Астрон. журн., **41** (1964).
53. *Ne'eman Y.* Astrophys. J., **141**, 1303 (1965).
54. *Oppenheimer J. R.*, *Synder H.*, Phys. Rev., **56**, 455 (1939).
55. *Misner C. W.*, *Thorne K. S.*, *Wheeler J. A.*, Gravitation. Freeman, San Francisco, 1973. [Имеется перевод: *Мизнер Ч.*, *Торн К.*, *Уилер Дж.* Гравитация.— М.: Мир, 1977.]
56. *Hawking S. W.*, Commun. Math. Phys., **43**, 199 (1975).
57. *Penrose R.*, In: Confrontation of Cosmological Theories with Observational Data. (IAU Symp. 63), ed. M. S. Longair. Reidel, Boston, 1974.
58. *Penrose R.*, In: Theoretical Principles in Astrophysics and Relativity, eds. N. R. Lebovitz, W. H. Reid and P. O. Vandervoort, University of Chicago Press, 1978.
59. *Penrose R.*, In: Proceedings of the First Marcel Grossmann Meeting on General Relativity, ICTP Trieste, ed. R. Ruffini, North-Holland, Amsterdam, 1977.
60. *Penrose R.*, In: Physics and Contemporary Needs, ed. Riazuddin. Plenum, New York, 1977.
61. *Penrose R.*, In: Perspectives in Geometry and Relativity, ed. B. Hoffmann. Indiana University Press, 1966.
62. *Parker L.*, In: Asymptotic Structure of Space-Time, eds. F. P. Esposito and L. Witten. Plenum. New York, 1977.
63. *Sexl R. U.*, *Urbantke H. K.*, Phys. Rev., **179**, 1247 (1969).
64. *Zel'dovich Ya. B.*, In: Gravitational Radiation and Gravitational Collapce (IAU Symp. 64) ed. C. M. DeWitt. Reidel, Boston, 1974.
65. *Eardley D. M.*, Phys. Rev. Lett, **33**, 442 (1974).
66. *Зельдович Я. Б.*, *Старобинский А. А.*, ЖЭТФ, **61**, 2161 (1971).
67. *Hawking S. W.*, Phys. Rev., **D13**, 191; **14**, 2460 (1976).
68. *Hawking S. W.*, Report to GR8 meeting, Waterloo, Ontario, 1977.
69. *Bekenstein J. D.*, Phys. Rev., **D7**, 2333 (1973); **9**, 3292 (1974).
70. *Penrose R.*, *Floyd R. M.*, Nature Phys. Sci., **229**, 177 (1971).
71. *Hawking S. W.*, Commun. Math. Phys., **25**, 152 (1972).
72. *Gibbons G. W.*, *Perry M. J.*, Proc. R. Soc. Lond., **A358**, 467 (1978).
73. *Wheeler J. A.*, Geometrodynamics. Academic Press, London, New York, 1962.
74. *Wald R. M.*, Commun. Math. Phys., **45**, 9 (1975).
75. *Parker L.*, Phys. Rev., **D12**, 1519 (1975).
76. *Pati J. C.*, *Salam A.*, Phys. Rev. Lett., **31**, 661 (1973); Phys. Rev., **D8**, 1240 (1973); **10**, 275 (1974).
77. *Sciama D. W.*, Modern Cosmology. Cambridge University Press, 1971.
78. *Weinberg S.*, Gravitation and Cosmology. Wiley, New York, 1972. [Имеется перевод: *Вейнберг С.* Гравитация и относительность.— М.: Мир, 1976.]
79. *Lynden-Bell D.*, *Lynden-Bell R. M.*, Mon. Not. R. Astron. Soc., **181**, 405 (1977).
80. *Белинский В. А.*, *Лифшиц Е. М.*, *Халатников И. М.*, УФН, **102**, 463 (1970).
81. *Kantowski R.*, *Sachs R. K.*, J. Math. Phys., **7**, 443 (1967).
82. *Gibbons G. W.*, *Hawking S. W.*, Phys. Rev., **D15**, 2752 (1977).

83. *Zel'dovich Ya. B.*, частное сообщение.
84. *Tolman R. C.*, Relativity, Thermodynamics and Cosmology. Clarendon Press, Oxford, 1934. [Имеется перевод: *Толмен Р.* Относительность, термодинамика и космология.— М.: Мир, 1974.]
85. *Israel W.*, J. Math. Phys., **4**, 1163 (1963).
86. *Liang E. P. T.*, Lett. Nuovo Cimento, **7**, ser. 2, 599 (1973).
87. *Simpson M., Penrose R.*, Int. J. Theor. Phys., **7**, 183 (1973).
88. *McNamara J. M.*, Proc. R. Soc. Lond., **A358**, 499 (1978).
89. *Geroch R., Kronheimer E. H., Penrose R.*, Proc. R. Soc. Lond., **A327**, 545 (1972).
90. *Penrose R.*, Proc. R. Soc. Lond., **A284**, 159 (1965).
91. *Bondi H.*, Cosmology, Cambridge University Press, 1952.
92. *Yodzis P., Seifert H.-J., Muller zum Hagen H.*, Commun. Math. Phys., **34**, 135 (1973); **37**, 29 (1974).
93. *Geroch R. P.*, J. Math. Phys., **11**, 437 (1970).
94. *Misner C. W.*, Astrophys. J., **151**, 431 (1968); Phys. Rev. Lett., **22**, 1071 (1969).
95. *Chitre D. M.*, Phys. Rev., **D6**, 3390 (1972).
96. *Дорошкевич А. Г., Новиков И. Д.*, Астрон. журн., **47**, 498 (1970).
97. *Лифшиц Е. М., Лифшиц И. М., Халатников И. М.*, ЖЭТФ, **59**, 322 (1970).
98. *MacCallum M. A. H.*, Nature, Phys. Sci., **230**, 112 (1971).
99. *Budic R., Sachs R. K.*, Gen. Relativ. Gravit., **7**, 21 (1976).
100. *Matzner R. A., Misner C. W.*, Astrophys. J., **171**, 415 (1972).
101. *Parker L.*, Nature, **261**, 20 (1976).
102. *Zel'dovich Ya. B.*, In: Magic Without Magic, John Archibald Wheeler. Freeman, San Francisco, 1972.
103. *Зельдович Я. Б., Новиков И. Д.* Релятивистская астрофизика.— М.: Наука, 1967.
104. *Barrow J. D., Matzner R. A.*, Mon. Not. R. Astron. Soc., **181**, 719 (1977).
105. *Barrow J. D.*, Nature, **267**, 117 (1977).
106. *Carter B.*, In: Confrontation of Cosmological Theories with Observational Data (IAU Symp. 63), ed. M. S. Longair, Reidel, Boston, 1974.
107. *Boltzmann L.*, Nature, **51**, 413 (1895).
108. *Partidge R. B., Wilkinson D. T.*, Phys. Rev. Lett., **18**, 557 (1967).
109. *Harrison E. R.*, Annu. Rev. Astron. Astrophys., **11**, 155 (1973).
110. *Carr B. J.*, Astrophys. J., **201**, 1 (1975).
111. *von Neumann J.*, Mathematical Foundations of Quantum Mechanics, Princeton University Press, 1955. [Имеется перевод: *фон Нейман Дж.* Математические основы квантовой механики.— М.: Наука, 1964.]
112. *Bialynicki-Birula I., Mycielski J.*, Ann. Phys. (N. Y.), **100**, 62 (1976).
113. *Lüders G.*, Ann. Phys., **2**, 1 (1957).
114. *Taylor J. C.*, Gauge Theories of Weak Interactions, Cambridge University Press.

VI. КВАНТОВАЯ ГРАВИТАЦИЯ: НОВЫЙ СИНТЕЗ

Б. С. Де Витт [1])

1. ВВЕДЕНИЕ

В своей речи, произнесенной в 1920 г. в Лейденском университете, Эйнштейн говорил об «изумительном упрощении теоретических принципов», достигнутом Г. А. Лоренцем путем «лишения эфира его механических качеств». Он описывал ситуацию в следующих словах:

«Согласно Лоренцу... один только эфир, а не вещество служит вместилищем электромагнитных полей... Одни лишь элементарные частицы способны к перемещениям, их электромагнитная активность целиком ограничивается тем, что они несут электрические заряды... Полушутя можно сказать (об эфире), что неподвижность — единственное механическое свойство, которого он не был лишен Лоренцем. К этому можно добавить, что единственное изменение, внесенное специальной теорией в концепцию эфира, состоит в лишении эфира последнего механического качества — его неподвижности [2]).

Эйнштейн указывал, что возможна такая точка зрения:

«Эфир вообще не существует. Электромагнитные поля являются не состояниями среды..., а такими же независимыми сущностями... не сводимыми к чему-либо еще... как атомы. Эта концепция подкрепляется тем обстоятельством, что электромагнитное излучение, подобно весомой материи, несет энергию и импульс, и, согласно специальной теории относительности, как вещество, так и излучение представляют собой лишь конкретные формы распределенной энергии».

Именно этого взгляда придерживались впоследствии многие теоретики. Сам же Эйнштейн сопротивлялся ему, ссылаясь на следующий простой пример:

«Представим себе волны на поверхности воды. Здесь мы можем описывать два совершенно различных явления. Либо мы можем наблюдать, как колеблющаяся поверхность, образующая границу между водой и воздухом, меняется во времени, либо с помощью маленьких поплавков мы можем наблюдать, как изменяются положения отдельных частиц воды. Если бы существование таких поплавков, следующих за движением частиц жидкости, было фундаментальным образом невозможно в физике, т. е. если бы нельзя было наблюдать ничего, кроме формы пространства, занимаемого водой... то у нас не было бы никаких оснований для предположения, что вода состоит из подвижных частиц. Но несмотря на это, мы могли бы характеризовать ее как некоторую среду».

Мысль Эйнштейна ясна. Согласно общей теории относительности, пространство-время само есть среда. «Отказаться от эфира,

---

[1]) *B. S. De Witt*, Department of Physics, The University of Texas at Austin, Austin, Texas, USA.

[2]) Немецкий оригинал можно найти в книге [31].

значит предположить, что пустое пространство не обладает никакими физическими свойствами». Общая теория относительности не только возвращает динамические свойства пустому пространству, но также приписывает ему энергию, импульс и момент количества движения. В принципе гравитационное излучение можно использовать как рабочее тело для реактивного двигателя. Поскольку гравитационные волны являются просто рябью на кривизне пространства-времени, «антиэфирист» должен был бы считать, что космический корабль, использующий такое рабочее тело, получает нечто за «ничто»: приобретает ускорение, просто выбрасывая один глубокий вакуум в другой. Этот пример не столь надуман, как кажется на первый взгляд. Нетрудно подсчитать, что звезда, испытывающая несимметричный (октупольный) коллапс, может получить таким способом приращение скорости около 100—200 $км \cdot с^{-1}$.

Однако концепция Эйнштейна идет дальше этого. Он считает, что кривизна — не единственное структурное свойство, которым обладает эфир. У эфира должны быть другие, более тонкие свойства, которые, подобно кривизне, лучше всего могут быть описаны на языке дифференциальной геометрии и дифференциальной топологии. Своей лейденской аудитории Эйнштейн так формулировал стоящие перед физикой задачи:

«Нам пока не известно, какая роль отведена этому новому эфиру в физике будущего. Мы знаем, что он определяет метрические отношения в пространственно-временно́м континууме... но мы не знаем, играет ли он существенную роль в структуре элементарных частиц... Было бы большим шагом вперед, если бы мы сумели объяснить гравитационное поле и электромагнитное поле как части единой структуры».

Эйнштейновский штурм проблемы единой теории поля и безуспешность этого штурма хорошо известны. Такая же неудача других ученых высочайшего уровня (достаточно назвать только трех из них: Вейль, Клейн, Паули) вызвала сильную реакцию среди теоретиков и отвратила их от подобных задач на многие годы. Все же эта мечта никогда не умирала совсем. Две черты эйнштейновской концепции никогда не теряли своей притягательности: потенциальное богатство реальности, основанной на геометрии, и предсказательная сила теорий, основанных на группах локальной инвариантности. Эти идеи пережили взрывоподобное начало золотой эры квантовой механики в 1926 г., великую эру квантовой электродиамики, последующее разочарование квантовой теорией поля и обращение с отчаяния к прикладному искусству дисперсионной теории.

В 60-е годы фундаментальная физика, в понимании Эйнштейна, начала оживать. Сначала интерес был сосредоточен на глобальных группах инвариантности, таких, как группа Пуанкаре, группы изотопического спина и $SU(3)$ в связи с экспериментальной физикой андронов. Однако к тому времени уже было известно о фундаментальном слабом взаимодействии и было открыто нарушение

четности. Поэтому адронные идеи стали применяться и к лептонам. Начала прокладывать путь алгебра токов, являя собой откровенный возврат к теоретико-полевым понятиям. Но в то же время наблюдалась активность иного рода.

В 1954 г. Янг и Миллс для того, чтобы превратить группу изотопического спина в локальную группу, ввели три новых векторных поля, взаимодействующих друг с другом и с материей способом, являющимся прямым обобщением взаимодействия в электродинамике. Очень скоро стало ясно, что идея Янга и Миллса может быть распространена на любую компактную группу Ли, и, таким образом, появился бесконечный ряд «неабелевых калибровочных теорий». Утияма [70] показал, что на эйнштейновскую теорию гравитации также можно смотреть как на неабелеву калибровочную теорию, причем главное различие состоит в том, что гравитационная калибровочная группа — группа диффеоморфизмов пространства-времени — имеет более богатую и более сложную структуру, чем группы Янга — Миллса. И обратно, неабелевы калибровочные теории могут быть выражены в геометрических, даже метрических, терминах (см. [25], задача 77).

Хотя в 50-е годы и в начале 60-х годов не было никаких экспериментальных данных о существовании полей Янга — Миллса, но гравитационное поле существовало, и это был упрямый факт. К сожалению, его рассматривали как некоторого рода аномалию природы во всех областях, кроме астрофизики. Лишь немногие одиночки, поддерживаемые верой в существенное единство природы, считали такую позицию неразумной. Они старались через квантование вернуть гравитации то центральное положение в физике, которое она занимала во времена Ньютона. В 1979 г., когда исполнилось сто лет со дня рождения Эйнштейна, стоит бросить общий взгляд на то, к чему ведут эти усилия, возбуждающие уже широкий интерес.

По крайней мере в течение десятилетия квантовая гравитация занималась почти исключительно формальными вопросами [1]). Первое большое продвижение, позволившее провести надежные вычисления чисто квантовых эффектов, было сделано лишь в 1962 г. Фейнман [39] показал в однопетлевом приближении, что наивные правила теории возмущений, используемые в квантовой электродинамике, будучи применены к квантовой гравитации и другим неабелевым теориям, нарушают унитарность. Чтобы поправить это, нужно добавить к каждому графу с замкнутой петлей компенсирующий граф, включающий замкнутую целоспиновую фермионную петлю. Прошло еще полдесятилетия, прежде чем был найден способ, как распространить рецепт Фейнмана на все порядки [23, 35].

[1]) Краткий итог работы, проделанной в этот начальный период, подведен в статье [22].

За это время были достигнуты большие успехи в теории слабых взаимодействий. Используя механизм, изобретенный Хиггсом [50] для введения нарушения симметрии, Вейнберг в 1967 г. и Салам в 1968 г. предложили то, что впоследствии развилось в первую успешную единую теорию поля. Это не та единая теория поля, о которой мечтал Эйнштейн в Лейдене в 1920 г., так как она объединяет электромагнетизм со слабым взаимодействием, а не с гравитацией. И все же она служит подтверждением геометрической в своей основе точки зрения Эйнштейна, так как центральную роль в ней играют локальные калибровочные группы.

Поскольку нарушение симметрии теории Вейнберга — Салама обусловлено исключительно вырождением вакуума (аналогичным вырождению основного состояния ферромагнетика), оно оставляет локальную калибровочную группу, которая является группой Янга — Миллса, в неприкосновенности. Использование этого факта позволило т'Хофгу [66] доказать, что эта теория перенормируема и, следовательно, дает вычислимые предсказания, которые могут быть экспериментально проверены. В своем доказательстве т'Хофт использовал правила вычислений, открытые при попытке квантовать гравитацию. Это заслуживает упоминания, поскольку здесь мы встречаемся с первым случаем в современной физике, когда исследование по теории тяготения дало непосредственный толчок в иной области физики. С тех пор теория гравитации быстро вернулась в основное русло физики.

В конце своей лейденской лекции Эйнштейн обратился с кратким предостережением: «Думая о ... будущем теоретической физики, мы не должны отвергать ту возможность, что факты, содержащиеся в квантовой теории, могут установить пределы теории поля, за которые он не может выйти». Эти слова были сказаны за восемь лет до рождения квантовой теории поля, поэтому Эйнштейн не мог знать, что квантовая теория не только не установит пределы теории поля, а наоборот, преобразует и обогатит ее. Все же он был прав, предполагая, что квантовая теория приведет к ряду новых проблем. Главной среди них является проблема неопределенностей, связанных с расходимостями, возникающими в ходе вычислений по теории возмущений; эта проблема особенно трудна в случае гравитационного поля. Эти неопределенности в квантовой гравитации на сегодняшний день не могут быть устранены на основе только той информации, которая (в принципе) может быть получена в конечном числе экспериментов. Иначе говоря, квантовая гравитация неперенормируема.

Некоторые аспекты данной проблемы рассматриваются в одном из разделов этой статьи. Однако чтобы видеть эту проблему в правильной перспективе, т. е. смотреть на нее с достаточным оптимизмом, необходимо сначала отметить те области, где общая теория относительности и квантовая теория поля в процессе нового синтеза уже взаимно обогатились.

2. КВАНТОВЫЙ ЭФИР

К наиболее впечатляющим примерам такого взаимообогащения следует отнести тот толчок, который был дан идеями общей теории относительности развитию понятия «вакуум», и ту поддержку, которую получила от квантовой теории поля идея, что на вакуум можно смотреть как на эфир, обладающий структурой. С первых дней возникновения квантовой электродинамики было известно, что напряженности полей в вакуумном состоянии испытывают случайные флуктуации, совершенно аналогичные нулевым колебаниям гармонического осциллятора, а когда учитывается взаимодействие с электронным полем, эти флуктуации сопровождаются рождением и аннигиляцией пар. Таким образом, вакуум оказывается состоянием постоянного беспорядочного движения.

С точки зрения Эйнштейна, было бы естественно считать, что флуктуации поля происходят в эфире и сообщают ему новые свойства в дополнение к геометрическим свойствам. Математическое описание вакуума, в котором эффективно воплощена эта идея, было дано уже довольно давно Швингером [64]. В присутствии внешнего источника квантованное поле, находившееся первоначально в вакуумном состоянии, не обязательно будет оставаться в этом состоянии. Швингер показал, что все физические свойства поля могут быть установлены, если знать, как с изменением источника варьирует амплитуда вероятности того, что поле останется в вакуумном состоянии. Функциональные производные вакуум-вакуумных амплитуд по источнику являются функциями отклика, которые описывают, как эфир реагирует на внешнее воздействие. Таким образом, в самом эфире содержится полный «проект» динамики поля.

Эфир можно зондировать не только источниками, но и иными средствами. Можно варьировать граничные условия и внешние поля. Например, вакуум-вакуумный матричный элемент тензора натяжений $T^{\mu\nu}$ для любой комбинации полей, включающей гравитационное поле, дается функциональной производной

$$\langle \text{out, vac} | T^{\mu\nu} | \text{in, vac} \rangle = -2i \frac{\delta}{\delta g_{\mu\nu}} \langle \text{out, vac} | \text{in, vac} \rangle. \qquad (1)$$

Здесь $|\text{in, vac}\rangle$ и $\langle\text{out, vac}|$ — начальный и конечный векторы вакуумного состояния соответственно, а $g_{\mu\nu}$— внешнее метрическое поле, часто называемое *фоновым полем*, которое служит произвольно выбранным началом отсчета квантовых флуктуаций гравитационного поля и может быть использовано для фиксации топологии пространственно-временного многообразия. В выражении (1) предполагается, что вакуумные состояния однозначно (хотя и не обязательно единственным образом) определены относительно данного фона и топологические переходы (если они вообще реально происходят) можно описать, лишь допуская переход по $g_{\mu\nu}$ в комплекс-

ную плоскость. Предполагается также, что любые расходимости, какие могут возникнуть, уже устранены перенормировками.

Аналогичное выражение в квантовой электродинамике имеет вид

$$\langle \mathrm{out, vac} | j^{\mu} | \mathrm{in, vac} \rangle = -i \frac{\delta}{\delta A_{\mu}} \langle \mathrm{out, vac} | \mathrm{in, vac} \rangle, \qquad (2)$$

где j^{μ} — вектор тока, а A_{μ} — вектор-потенциал внешнего или фонового электромагнитного поля. Выражение (2), вообще говоря, отлично от нуля всюду, где отлично от нуля фоновое поле; это явление известно как *поляризация вакуума*. Подобным же образом выражение (1) отлично от нуля, вообще говоря, всюду, где фоновая геометрия искривлена. Однако кривизна — не единственный источник гравитационной «поляризации вакуума». *Вносит вклад также топология*. Это означает, что свойства эфира зависят от многообразия в целом!

Этот факт столь поразителен, что заслуживает более подробного рассмотрения. Такое явление впервые было открыто в эффекте Казимира. При вычислении сил Ван-дер-Ваальса между очень близкими молекулами Казимир [13] обнаружил, что энергию взаимодействия можно представить как сумму членов, зависящих кроме межмолекулярных расстояний и внутримолекулярных параметров также и от кривизны молекулярной поверхности. Но один член не зависит ни от кривизны, ни от особенностей молекул. Наличие этого члена говорит о том, что между двумя плоскопараллельными проводящими пластинами в вакууме должна существовать сила притяжения. Этот эффект вскоре был подтвержден в лабораториях фирмы «Филипс». Поскольку сила притяжения крайне мала, необходимо было принять исключительные меры предосторожности, чтобы поверхности были абсолютно чисты, нейтральны и микроскопически плоски; только при этих условиях их можно довести почти до соприкосновения без возникновения каких-либо посторонних эффектов.

За эффект Казимира ответственно электромагнитное поле, а многообразием при этом является плоскопараллельный *зазор* между проводящими пластинами. По математической терминологии, это — *неполное* многообразие [1]). Однако ничто не препятствует построению физики в неполном многообразии, если наложены подходящие граничные (краевые) условия, в данном случае условие идеальной проводимости на границах. Свойства эфира между проводниками целиком определяются полевыми функциями Грина (функциями отклика), соответствующими зазору.

Рассмотрим сначала бесконечный вакуум в пространстве Минковского, т. е. стандартный вакуум физики элементарных частиц. Оператор $\mathbf{T}^{\mu\nu}$, описывающий энергию, импульс и натяжения в

---

[1]) В русской математической литературе более употребителен термин «многообразие с краем».— *Прим. перев.*

электромагнитном поле, формально является билинейным произведением операторнозначных обобщенных функций (полевых операторов) и поэтому не имеет смысла. Ему придается смысл с помощью вычитательной процедуры, по которой ожидаемое значение $\langle T^{\mu\nu}\rangle$ полагается равным нулю в вакууме в пространстве Минковского. Это вычитание соответствует пренебрежению нулевой энергией полевых осцилляторов, но оно ни в коем случае не является произвольным. $\langle T^{\mu\nu}\rangle$ *должно* обращаться в нуль в пустом пространстве-времени Минковского, если квантовая теория поля должна быть полностью согласована с общей теорией относительности. Вакуум пространства Минковского служит тем стандартом, с которым должны сопоставляться любые другие возможные вакуумы.

Допустим теперь, что в вакуум Минковского внесен одиночный бесконечный плоский проводник. Можно представить себе, что проводник внесен адиабатически из бесконечности, так что поле не испытывает никакого возбуждения и остается в своем основном состоянии. Интересующее нас многообразие становится при этом бесконечным полупространством. Введем координаты Минковского x^μ, $\mu=0, 1, 2, 3$, ориентированные так, что ось x^3 перпендикулярна плоскости проводника. Из соображений симметрии ясно, что $\langle T^{\mu\nu}\rangle$ должно быть диагональным и независящим от x^0, x^1, x^2. Более того, поскольку идеальный проводник остается идеальным проводником в любом состоянии движения параллельно своей поверхности, вакуумные натяжения вблизи него должны иметь один и тот же вид независимо от того, насколько быстро наблюдатель скользит вдоль этой поверхности. Иными словами, эфир всегда сохраняет свои релятивистские свойства, и, следовательно, $\langle T^{\mu\nu}\rangle$ должно быть инвариантным относительно преобразований Лоренца, соответствующих бустам, параллельным плоскости (x^1, x^2). Это означает, что первые три столбца и строки $\langle T^{\mu\nu}\rangle$ должны быть пропорциональны метрическому тензору (2+1)-мерного пространства Минковского, а именно diag (—1, 1, 1). Если к этому выводу добавить еще то наблюдение, что в случае электромагнитного поля $T^\mu_\mu=0$, то мы придем к заключению, что $\langle T^{\mu\nu}\rangle$ имеет вид

$$\langle T^{\mu\nu}\rangle = f(x^3)\times \mathrm{diag}(-1, 1, 1, -3). \tag{3}$$

Но это еще не все. Можно установить также и вид функции $f(x^3)$. Для этого нужно прибегнуть к закону сохранения $\langle T^{\mu\nu}{}_{,\nu}\rangle=\langle T^{\mu\nu}\rangle_{,\nu}=0$. В частности,

$$0=\langle T^{3\nu}\rangle_{,\nu}=-3f'(x^3), \tag{4}$$

откуда следует, что f есть постоянная, не зависящая от x^3. Далее, $\langle T^{\mu\nu}\rangle$ имеет размерность энергии. Единственные фундаментальные постоянные, которые входят в теорию,— это $\hbar$ и c. Чтобы получилась постоянная, имеющая размерность энергии, необходима единица длины, массы или времени. В данной задаче нет никаких

величин с такими размерностями. Поэтому можно прийти к единственному заключению, что $f=0$ и, следовательно, в бесконечном полупространстве $\langle \mathbf{T}^{\mu\nu} \rangle = 0$.

Этот вывод подтверждается прямым вычислением. Функцию Грина для бесконечного полупространства легко построить из функции Грина для пространства Минковского методом изображений. Тогда $\langle \mathbf{T}^{\mu\nu} \rangle$ получается соответствующим дифференцированием полученной функции Грина и последующим сведéнием вместе пространственно-временных точек, от которых она зависит. Результат, конечно, расходится и должен быть перенормирован путем вычитания соответствующего результата для пространства Минковского. Это эквивалентно вычитанию функции Грина для пространства Минковского из функции Грина для полупространства. Хотя результирующая перенормированная функция Грина сама не обращается в нуль, она дает $\langle \mathbf{T}^{\mu\nu} \rangle = 0$.

Все приведенные соображения относительно вида $\langle \mathbf{T}^{\mu\nu} \rangle$ столь же верны и для многообразия-зазора, за тем исключением, что теперь имеется естественная единица длины — расстояние a, разделяющее параллельные проводящие плоскости. Поэтому в области между проводниками следует ожидать, что

$$\langle \mathbf{T}^{\mu\nu} \rangle = f(a) \times \mathrm{diag}(-1,\ 1,\ 1,\ -3). \qquad (5)$$

Вид функции $f(a)$ может быть определен из рассмотрения работы необходимой для адиабатического разведения проводников. Из анализа случая полупространства мы знаем, что на проводники не действуют извне никакие силы. Но есть внутренняя сила, стремящаяся свести их вместе, и ее величина на единицу площади равна $3f(a)$. Если проводники раздвигаются на расстояние da, то для этого на единицу площади нужно совершить работу $dW = 3f(a)da$. Она должна проявиться как прирост энергии на единицу площади, равный $E = -af(a)$. Положив $dW = dE$ и интегрируя, немедленно получаем

$$f(a) = \frac{A}{a^4}, \qquad (6)$$

где A — некоторая универсальная константа.

Вид выражения (6) можно также вывести из анализа размерностей. Единственная комбинация из $\hbar$, c и a, дающая плотность энергии, есть $\hbar c/a^4$. В дальнейшем мы положим $\hbar = c = 1$. Тогда постоянная A будет безразмерным числом.

Чтобы оценить значение A, необходимо прямое вычисление. Вновь методом изображений можно построить функцию Грина и, вычитая из нее функцию Грина для пространства Минковского, получить перенормированную функцию Грина. Перенормированное $\langle \mathbf{T}^{\mu\nu} \rangle$ уже не обращается в нуль. При этом подтверждаются выражения (5) и (6), и для A получается значение $\pi^2/720$. В преде-

лах ожидаемых ошибок это значение находится в согласии с опытом.

Можно видеть, что плотность энергии эфира между проводниками *отрицательна*. Эта энергия ничтожно мала, на много порядков меньше энергии, необходимой для создания гравитационного поля, которое можно было бы измерить. Но легко придумать мысленные эксперименты, в которых нарушается закон сохранения энергии, если эта энергия не будет включена в источник гравитационного поля. Плотность энергии в квантовом эфире часто оказывается отрицательной. Поэтому квантовая теория нарушает посылки знаменитых теорем Пенроуза — Хокинга относительно неизбежности сингулярностей в пространстве-времени, которые в конечном счете предсказывают крах классической общей теории относительности.

2.1. ДРУГИЕ ТОПОЛОГИИ

Для того чтобы $\langle T^{\mu\nu}\rangle$ было отлично от нуля в плоском пространстве, вовсе не обязательно, чтобы, как в случае эффекта Казимира, многообразие было неполным. В полных многообразиях также может наблюдаться подобное явление; например, в многообразиях $R\times\Sigma$, где «слои» Σ — плоские, пространственноподобные гиперповерхности Коши, имеющие одну из следующих топологий: $R^2\times S^1$, $R\times T^2$, T^3, $R\times K^2$ и т. д., где T^n есть n-мерный тор, K^2— двумерная бутылка Клейна и т. д. Случай $\Sigma=R^2\times S^1$ имеет наиболее близкое сходство с эффектом Казимира. Единственное отличие состоит в том, что вместо наложения граничных условий о том, что поверхности, ограничивающие зазор, образованы проводником, теперь накладываются периодические граничные условия. $\langle T^{\mu\nu}\rangle$ снова принимает вид (5), (6), где теперь a — период координаты x^3. Снова нетрудно подсчитать функцию Грина, и в этом случае мы получаем $A=\pi^2/45$. Случаи $\Sigma=R\times T^2$, $\Sigma=T^3$ более сложны. Хотя $\langle T^{\mu\nu}\rangle$ не зависит от координат, его вид теперь уже не дается выражениями (5) и (6), а зависит от отношений периодичностей по разным координатам. В случае $\Sigma=R\times K^2$ $\langle T^{\mu\nu}\rangle$ перестает быть независимым от координат и становится периодическим (и гладким) [1].

Одним из преимуществ изучения квантовой теории поля на полных многообразиях является то, что можно рассматривать любое поле. Не нужно заботиться о том, каковы граничные условия, аналогичные граничным условиям на проводнике для электромагнитного поля. Можно ввести скалярное, спинорное и даже гравитационное поле. Спинорные поля представляют особый интерес из-за того, что на некоторых многообразиях, например на $\Sigma=$

[1]) Автор благодарит Ч. Харта за информацию об этих случаях, в каждом из которых плотность энергии отрицательна.

$=R^2 \times S^1$, можно ввести спинорные поля, которые гомотопически не эквивалентны. Это означает, что можно ввести более одного «вакуумного» состояния.

Ситуации такого рода изучались в связи с «кинками», солитонами и инстантонами (по этому вопросу см. обзор [56]). Общая теория относительности с тем богатством топологий, которые она допускает, увеличивает разнообразие и сложность этих ситуаций. Более того, в качестве модели других (обычно более простых) полевых теорий она привлекла внимание к факту, который до этого часто не замечали, а именно к тому, что конфигурационное пространство любого набора взаимодействующих полей само является римановым многообразием или в более общем случае (когда участвуют фермионные поля) градуированным римановым многообразием [1]), обладающим метрикой, которая определяется (по крайне мере частично) лагранжианом полей. Топология этого многообразия вовсе не обязательно будет тривиальной [2]).

Это конфигурационное пространство, не жертвуя его римановым характером, можно также рассматривать как (градуированное) расслоенное пространство над пространством-временем. Топология этого расслоенного пространства не обязательно тривиальна. Примером может служить гомотопическая неэквивалентность классов спинорных полей на $\Sigma=R^2 \times S^1$. Более простой пример на $R^2 \times S^1$ был предложен Ишэмом [53]. Рассмотрим нейтральное скалярное поле. Классически оно является отображением из данного пространства-времени на действительную прямую R. Поэтому его конфигурационное пространство есть расслоенное пространство, где слоями являются копии R. Предположим, что расслоение «скручено» так, что поле, вместо того чтобы быть периодическим над $\Sigma=R^2 \times S^1$, становится антипериодическим. Эта антипериодичность должна отразиться в функции Грина, а следовательно, и вакуумный тензор натяжений тоже должен измениться. Прямым вычислением для безмассового поля можно получить, что $\langle \mathbf{T}^{\mu\nu} \rangle$ по-прежнему сохраняет вид (5) и (6), но теперь $A=-7\pi^2/720$. В этом случае плотность энергии *положительна*. Когда расслоение не «скручено», плотность энергии отрицательна, как это было в случае электромагнитного поля.

Во всех приведенных выше примерах фоновое многообразие — плоское. Энергия и натяжение в эфире целиком обусловлены топологией. Не всякая топология допускает плоскую метрику, примерами могут служить $\Sigma=S^3$ и $\Sigma=R \times S^2$. В этих случаях также $\langle \mathbf{T}^{\mu\nu} \rangle$ отлично от нуля. Однако «поляризация вакуума» теперь

---

[1]) Введением в градуированные многообразия может служить работа [59].

[2]) Например, в случае поля синус-Гордона. Другим интересным примером является само гравитационное поле; описание его конфигурационного пространства как «слоистого» бесконечномерного риманова многообразия см. в работе [40]. Исследование некоторых топологических свойств этого многообразия проведено в работах [22, 24].

вызвана не одной только топологией; кривизна также играет определенную роль. Кривизна к тому же усложняет алгоритм перенормирования и приводит к явлению *аномалий следа* (trace anomalies): формальное равенство $T^{\mu}_{\mu}=0$, справедливое для конформно-инвариантных классических теорий поля, для квантованных полей при наличии кривизны перестает выполняться [1]). Краткое обсуждение этих проблем дано в одном из последующих разделов. Здесь мы лишь отметим, что аномалии следа во многих отношениях подобны *аномалиям аксиально-векторного тока* теорий слабых взаимодействий, в которых также перестает выполняться некоторое формальное равенство (условие на дивергенцию).

Аномалии аксиально-векторного тока играют определенную роль в теории π^0-распада и при анализе ограничений, которые должны быть наложены на единые теории поля типа теории Вейнберга — Салама. Они были известны в течение многих лет, и первоначально никто не думал, что они имеют какое-либо отношение к общей теории относительности. Однако Кимура [58] обнаружил, что в действительности они тесно связаны с топологией пространства-времени. Этот факт и его связь с числом Понтрягина и индексом Атья — Зингера [2, 3] получили дальнейшее подтверждение в работах Дельбурго и Салама [17], Эгучи и Фройнда [30] и Джэкива и Ребби [57]. Его возможная связь с квантовомеханическим туннелированием была отмечена Хокингом [47, 48] на основе работ т'Хофта [67, 68], где рассматривается фоновое поле Янга — Миллса. Пока еще рано настаивать на большом значении этих достижений, но ясно, что они вместе со всеми другими упомянутыми выше примерами гарантируют, что топология непременно будет играть важную роль в квантовой теории поля будущего.

2.2. ГРАНИЦЫ С КРИВИЗНОЙ; УСКОРЕНИЕ

При анализе эффекта Казимира были введены два неполных многообразия — полупространство и зазор. Границы (края) этих многообразий плоские. Что получится, если границы искривлены? Рассмотрим сначала случай, когда граница состоит из двух непараллельных плоских проводников, соединенных по линии пересечения. Можно считать, что кривизна сосредоточена на этой линии. Как всегда, для построения функции Грина в нашем распоряжении имеется метод изображений. Оказывается, что перенормированное $\langle T^{\mu\nu}\rangle$ зависит от угла пересечения и изменяется как четвертая степень обратного расстояния от линии пересечения [28]. В более общем случае $\langle T^{\mu\nu}\rangle$ имеет такое поведение вблизи линии пересечения каждого двухгранного угла любой многогранной граничной поверхности.

[1]) Аномалии следа открыли Каппер и Дафф [12]. Но общий их вид не был известен до работ [15, 18, 27]. Точные значения коэффициентов были вычислены в работах [8, 10, 14].

Поверхность искривленного проводника может рассматриваться как предел последовательности многогранных поверхностей, и $\langle T^{\mu\nu} \rangle$ может быть аппроксимировано путем перехода к такому пределу. Дойч [1]) сумел показать, что вблизи проводника $\langle T^{\mu\nu} \rangle$ пропорционально сумме двух обратных главных радиусов кривизны и изменяется по закону обратного куба расстояния от проводника. По мере приближения к проводнику плотность энергии в эфире стремится к $-\infty$ на вогнутой стороне и к $+\infty$ на выпуклой стороне.

Изменение по закону обратного куба расстояния фактически говорит о том, что приближение идеального проводника теряет силу. Из этого закона следует не только обращение в бесконечность ($-\infty$ на вогнутой стороне и $+\infty$ на выпуклой стороне) энергии в эфире на единицу площади проводника, но также и бесконечные значения испытываемых проводником напряжений. Проводник не остается идеальным при произвольно высоких частотах фотонов. Немедленно вступает в игру эффективное энергетическое обрезание или глубина скин-слоя, зависящие от конкретного атомного строения проводника. Это согласуется с результатами Казимира, полученными при изучении сил Ван-дер-Ваальса. Те члены в энергии взаимодействия, которые зависят от кривизны молекулярной поверхности, зависят также от конкретного молекулярного строения, т. е. от таких фундаментальных постоянных, как e и m.

В приведенных примерах предполагается, что проводящие поверхности покоятся или движутся равномерно. Поэтому кривизна границы в соответствующих неполных многообразиях является чисто пространственной. Если проводнику сообщено ускорение, то граница многообразия обладает также *кривизной во времени*. Особенно простым примером может служить случай, когда бесконечному плоскому проводнику сообщается ускорение, перпендикулярное его поверхности. Если ускорение меняется со временем, то проводник, вообще говоря, излучает или поглощает фотоны, т. е. обменивается энергией с эфиром, но если ускорение постоянно, то на вогнутой стороне поверхности может установиться равновесие (ускорение происходит *в направлении* неполного многообразия). Причина здесь в том, что на вогнутой стороне может быть введено времениподобное векторное поле Киллинга, по отношению к которому проводник покоится. Во всех предыдущих примерах существует по крайней мере одно *геодезическое* времениподобное векторное поле Киллинга, и неявно подразумевается, что «вакуум» определен относительно этого поля через стандартное разложение на положительные и отрицательные частоты. Но геодезический характер этого поля не существен. Разбиение на положительно- и отрицательно-частотные части и соответствующее определение вакуума можно осуществить относительно любого

---

[1]) Частное сообщение.

времениподобного векторного поля Киллинга независимо от того, геодезическое оно или нет.

В случае плоского проводника, которому сообщено постоянное ускорение, этот вакуум обладает некоторыми замечательными свойствами. Пусть a_0 — ускорение проводника и p — пространственно-временная точка на расстоянии s от нее. Кандела и Дойч [11] показали, что для значений s, малых по сравнению с a_0^{-1}, перенормированный тензор натяжений $\langle T^{\mu\nu}\rangle$ в точке p, рассматриваемый в локальной системе покоя относительно векторного поля Киллинга, меняется как $1/s^3$, т. е. точно так же, как вблизи неускоренного искривленного проводника. Действительно, если a_0^{-1} играет роль радиуса кривизны, то формулы для обоих этих случаев одинаковы: совпадает все — и числовые коэффициенты, и знаки. В частности, локальная плотность энергии отрицательна.

Вид, который принимает $\langle T^{\mu\nu}\rangle$ при $s\gg a_0^{-1}$, еще более примечателен. Он совпадает, но *с обратным знаком* с тем, что было бы, если бы эфир представлял собой фотонный газ в тепловом равновесии. Локальная температура меняется стандартным релятивистским образом: как обратная длина локального вектора Киллинга. В рассматриваемом случае это означает, что температура пропорциональна локальному ускорению. В единицах, в которых $k=1$, коэффициент пропорциональности оказывается равным $1/2\pi$.

Математический анализ этой системы упрощается, если ускорение a_0 устремить к бесконечности. При этом удобно ввести «координаты Риндлера» τ, ζ, связанные с координатами Минковского x^0, x^3 равенствами [63]

$$x^0=a^{-1}e^{a\zeta}\,\mathrm{sh}\,a\tau,\qquad x^3=a^{-1}e^{a\zeta}\,\mathrm{ch}\,a\tau. \tag{7}$$

Здесь ускорение (вектора Киллинга $\partial/\partial\tau$) происходит в направлении оси x^3 и имеет локальное значение a на гиперповерхности $\zeta=0$ и $ae^{-a\zeta}$ повсюду вне ее. Проводник перемещен к краю многообразия $\zeta=-\infty$ (где его присутствие фактически несущественно). Это многообразие известно как «клин» Риндлера и охватывает область $x^3>|x^0|$.

Локальная температура эквивалентного фотонного газа по всему клину Риндлера теперь дается формулой $T=ae^{-a\zeta}/2\pi$. Это означает, что в локальной системе покоя относительно $\partial/\partial\tau$ тензор натяжений меняется как четвертая степень обратного расстояния от острия клина. Это можно сравнить с поведением $\langle T^{\mu\nu}\rangle$ вблизи линии пересечения двух проводящих плоскостей. Но теперь мы сталкиваемся с понятием температуры смысл которого еще нужно уяснить. Кроме того, необходимо ответить также на следующий вопрос: как нужно понимать *отрицательность* плотности энергии?

Дело выглядит так, словно основное состояние клина Риндлера (так называемый вакуум Риндлера) каким-то образом находится *ниже* абсолютного нуля температуры! Для того чтобы привести энергию состояния к энергии вакуума Минковского, необходимо

добавить фотоны, и эти фотоны должны иметь тепловое распределение.

Как ни удивительна такая интерпретация, она все-таки верна. Этот и многие другие примеры, найденные за последние несколько лет, привели к важным изменениям в наших представлениях о «частицах» и о способах определения их и связанных с ними вакуумных состояний. Как это было ранее с теорией относительности и с квантовой теорией, мы оказались вынужденными обратиться к операционным определениям. В нашем случае мы должны задаться вопросом: каким был бы отклик детектора частиц в данной ситуации? [1])

Чтобы увидеть, как ответ на этот вопрос разрешает загадку с отрицательной энергией, рассмотрим сначала детектор частиц в обычном пространстве-времени Минковского (без проводников). Пусть для простоты электромагнитное поле заменено безмассовым скалярным полем ϕ, и пусть взаимодействие между детектором и полем простого монопольного типа и описывается лагранжианом вида

$$\mathrm{L}_{\text{вз}} = \mathrm{m}(\tau)\,\phi\,(x(\tau)). \tag{8}$$

Здесь функции $x^\mu(\tau)$ определяют мировую линию детектора (идеализируемого в виде точечного объекта), а оператор $\mathrm{m}(\tau)$ определяет его монопольный момент в момент собственного времени τ. Допустим, что детектор обладает набором внутренних энергетических собственных состояний, описываемых векторами $|E\rangle$, причем $E=0$ соответствует основному состоянию. Отклик детектора тогда будет связан с матричными элементами

$$\langle E\,|\,\mathrm{m}(\tau)\,|\,E'\rangle = \langle E\,|\,\mathrm{m}(0)\,|\,E'\rangle\, e^{i(E-E')\tau}. \tag{9}$$

Предположим, что детектор вначале находится в своем основном состоянии, а поле — в состоянии, описываемом символом Ψ. Тогда амплитуда того, что эта сложная система перейдет из состояния $(0, \Psi)$ в состояние (E, Ψ'), определяется формулой

$$A(E, \Psi'\,|\,0, \Psi) = \left\langle E, \Psi' \middle| T\left(\exp\left(i\int_{-\infty}^{\infty} \mathrm{L}_{\text{вз}}d\tau\right)\right)\middle| 0, \Psi\right\rangle, \tag{10}$$

где T — оператор хронологического упорядочения. Если $E>0$ и (перенормированный) монопольный момент настолько мал, что радиационными поправками к лагранжиану взаимодействия можно пренебречь, то эта амплитуда адекватно определяется первым по-

[1]) *В. Унрю* [69] первый предпринял систематическое исследование этого вопроса.

рядком теории возмущений:

$$A(E,\ \Psi'\,|\,0,\ \Psi) \approx i\left\langle E,\ \Psi'\,\middle|\int\limits_{-\infty}^{\infty} \mathrm{m}(\tau)\,\phi(x(\tau))\,d\tau\,\middle|\,0,\Psi\right\rangle =$$

$$= i\,\langle E\,|\,\mathrm{m}(0)\,|\,0\rangle \int\limits_{-\infty}^{\infty} e^{iE\tau}\,\langle\Psi'\,|\,\phi(x(\tau))\,\Psi\rangle\,d\tau. \qquad (11)$$

Полная вероятность того, что детектор перейдет в возбужденное состояние с энергией E, тогда равна

$$P(E) = \sum_{\Psi'} |\,A(E,\ \Psi'\,|\,0,\ \Psi)|^2 =$$

$$= |\langle E\,|\,\mathrm{m}(0)\,|\,0\rangle\,|^2 \int\limits_{-\infty}^{\infty} d\tau \int\limits_{-\infty}^{\infty} d\tau' e^{-iE(\tau-\tau')} \langle\Psi\,|\,\phi(x(\tau))\,\phi(x(\tau'))\,|\,\Psi\rangle. \quad (12)$$

Мы видим, что отклик детектора зависит от матричного элемента монопольного момента и от фурье-образа (вдоль мировой линии) функции Вайтмана для поля.

Если поле находится в стандартном вакуумном состоянии пространства Минковского, а детектор движется вдоль геодезической мировой линии, то в функции Вайтмана имеются только положительные частоты, фурье-образ ее в формуле (12) равен нулю и детектор остается в своем основном состоянии ($P(E)=0$). Если детектор испытывает ускорение, то $P(E)$, вообще говоря, уже не равно нулю; статистически предсказуемым образом происходят переходы в возбужденные состояния.

Обычно эти переходы не рассматриваются как сигналы о поглощении детектором фотонов, поскольку нет никаких фотонов (пространства Минковского), которые могли бы быть поглощены. Наоборот, считается, что детектор *излучает* фотоны. Если бы детектор был инертным, не обладал внутренними степенями свободы, то схема излучения фотона была бы такая же, как для заданного ускоренного источника. При наличии внутренних степеней свободы эта схема меняется: иногда излученный фотон мягче, чем он был бы при отсутствии у детектора внутренних степеней свободы. Детектор забирает часть его энергии и переходит в возбужденное состояние.

Когда детектор испытывает постоянное ускорение, возможна альтернативная точка зрения. Если ускорение равно a, то можно считать, что детектор движется вдоль линии $x^1=\mathrm{const}$, $x^3=\mathrm{const}$, $\zeta=0$ в системе координат Риндлера (7). Граница клина Риндлера является горизонтом для детектора. Для детектора этот клин является вселенной. В клине поле может быть в риндлеровском вакуумном состоянии. Тогда функция Вайтмана (12) содержит лишь положительные частоты *по отношению к риндлеровскому времени* τ. Какие-либо «риндлеровские фотоны» отсутствуют, и детектор остается в основном состоянии.

Очевидно, что риндлеровские фотоны и фотоны пространства Минковского — не одно и то же. Равноускоренный детектор регистрирует риндлеровские фотоны, неускоренный — фотоны пространства Минковского. Оба вида фотонов вносят вклад в $\langle \mathbf{T}^{\mu\nu} \rangle$, но они относятся к разным основным состояниям. Вакуум пространства Минковского заполнен риндлеровскими фотонами, хотя он лишен фотонов пространства Минковского. Энергия, которую несут эти риндлеровские фотоны, поднимает $\langle \mathbf{T}^{00} \rangle$ с отрицательного значения в риндлеровском вакууме к нулевому значению в вакууме Минковского.

Тепловой (отрицательный) характер $\langle \mathbf{T}^{\mu\nu} \rangle$ в риндлеровском вакууме можно понять, исходя из того факта, что функция распределения риндлеровских фотонов в вакууме Минковского тепловая. Это можно продемонстрировать несколькими способами. Один способ — вычислить функцию Вайтмана (12) для случая равноускоренного движения в вакууме Минковского. Результат оказывается равным

$$\langle \Psi | \phi(x(\tau))\, \phi(x(\tau')) | \Psi \rangle = -\frac{(a/2\pi)^2}{4\,\mathrm{sh}^2 \frac{1}{2} a\,(\tau - \tau' - i0)} . \qquad (13)$$

По своей форме это функция Вайтмана для чистого состояния, однако можно показать, что она идентична функции Вайтмана для смешанного состояния $\mathrm{tr}\,[\rho\phi(x(\tau))\,\phi(x(\tau'))]$, где ρ — матрица плотности, описывающая термостат из риндлеровских фотонов при температуре $T = a/2\pi$ относительно мировых линий $\zeta = 0$:

$$\rho = \frac{e^{-2\pi \mathbf{H}/a}}{\mathrm{tr}\,(e^{-2\pi \mathbf{H}/a})} , \qquad (14)$$

где $\mathbf{H}$ — риндлеровский гамильтониан относительно мировых линий $\zeta = 0$.

Другой способ продемонстрировать тепловой характер распределения риндлеровских фотонов, способ, который объясняет, каким образом чистое состояние может выглядеть как смешанное, состоит в том, чтобы ввести полный набор функций — риндлеровских мод не только в исходном клине Риндлера, но также и во всех остальных трех квадрантах ($x^0 > |x^3|$, $x^3 < -|x^0|$ и $x^0 < -|x^3|$). Этот набор может быть использован для того, чтобы определить обобщенный риндлеровский вакуум и соответствующие риндлеровские фотоны во всех четырех квадрантах. Затем можно вычислить боголюбовские коэффициенты, связывающие риндлеровские базисные функции и базисные функции в пространстве Минковского, что позволяет непосредственно вычислить функцию распределения риндлеровских фотонов в вакууме Минковского во всех четырех квадрантах. Для боголюбовского разложения характерно, что каждый вакуум представим как некоторое распределение статистически независимых *пар* фотонов по отношению к другому вакууму. Оказы-

вается, что в данном случае, кроме того, что они имеют тепловое распределение, составляющие каждой риндлеровской пары соответствуют базисным функциям, имеющим непересекающиеся носители в пространстве-времени. Две составляющие одной пары никогда не находятся в одном и том же квадранте. Поэтому все риндлеровские фотоны в исходном клине Риндлера статистически независимы (некогерентны), а это означает, что вакуум Минковского локально идентичен риндлеровскому тепловому состоянию.

2.3. ЧЕРНЫЕ ДЫРЫ

Тепловые состояния играют особенно важную роль в теории черных дыр. Легче всего это увидеть в случае геометрии Шварцшильда. Пусть t — стандартная шварцшильдовская временна́я координата. Тогда $\partial/\partial t$ — времениподобный вектор Киллинга повсюду вне горизонта, и вакуумное состояние определяется по отношению к нему. На больших расстояниях от черной дыры этот вакуум неотличим от обычного вакуума Минковского. В частности, в этой области $\langle \mathbf{T}^{\mu\nu} \rangle$ (перенормированное) обращается в нуль. Вблизи горизонта, напротив, этот вакуум приобретает многие свойства риндлеровского вакуума: детектор частиц, покоящийся относительно $\partial/\partial t$, остается в своем основном состоянии. Более того, в локально лоренцевой системе отсчета на горизонте $\langle \mathbf{T}^{00} \rangle$ приобретает бесконечное отрицательное значение, точно так же как на границе клина Риндлера в риндлеровском вакууме.

Имеется другой «вакуум», который совместим с геометрией Шварцшильда,— состояние, для которого $\langle \mathbf{T}^{\mu\nu} \rangle$ (в локальной системе отсчета) остается на горизонте конечным. Это состояние фиксируется требованием, чтобы свободно падающий детектор не испытывал никаких стимулированных переходов вблизи горизонта. В этой области такое состояние, очевидно, похоже на вакуум Минковского. На бесконечности, напротив, оно имеет тепловой характер, что можно установить из следующих рассуждений.

Пусть M — масса черной дыры. Выберем единицы, в которых гравитационная постоянная G равна единице. Пусть детектор находится в покое относительно $\partial/\partial t$ и расположен при $r=2M+\varepsilon$, $\varepsilon \gg 2M$, где r — обычная шварцшильдовская радиальная координата. Для того чтобы это положение детектора не менялось, он должен испытывать абсолютное ускорение, равное $(2M/\varepsilon)^{1/2}/4M$. Поскольку это состояние обладает локальными свойствами вакуума Минковского вблизи $r=2M$, отсюда следует, что детектор должен реагировать, по крайней мере в области высоких частот, так, как если бы он был помещен в фотонный термостат с температурой $T_\varepsilon = (2M/\varepsilon)^{1/2}/8\pi M$. Фотоны в этом термостате соответствуют базисным функциям, основанным на $\partial/\partial t$. Они реальны, поскольку они несут энергию, увеличивающую $\langle \mathbf{T}^{00} \rangle$ на горизонте от $-\infty$ до конечного значения. Более того, они могут уйти на бесконечность, где вслед-

ствие фактора красного смещения $(\varepsilon/2M)^{1/2}$ у них будет температура

$$T = \frac{1}{8\pi M}\,. \tag{15}$$

Это знаменитая температура, введенная Хокингом [46]. Он впервые встретился с ней при изучении черных дыр, образующихся при коллапсе. Метрика коллапса допускает корректную постановку задачи Коши для базисных функций поля излучения. Следовательно, если $\langle \mathbf{T}^{\mu\nu} \rangle$ первоначально несингулярно, оно должно оставаться несингулярным до достижения геометрической сингулярности при $r=0$. В частности, оно должно быть гладким на горизонте. Предполагалось, что черная дыра, о которой говорилось до сих пор, — это «вечная» черная дыра, горизонт которой состоит как из будущей, так и из прошлой частей. Черная дыра, образующаяся при коллапсе, имеет только горизонт будущего. Условие, что $\langle \mathbf{T}^{\mu\nu} \rangle$ должно быть гладким на горизонте, ведет (начиная с предколлапсного состояния, в котором на бесконечности нет частиц) к конечному состоянию, в котором фотоны на бесконечности тепловые с $T=1/8\pi M$, но только вылетающие. В случае вечной черной дыры состояние на бесконечности есть состояние термостата, находящегося в равновесии с черной дырой; при этом поглощается столько же фотонов, сколько излучается.

Введение Хокингом соответствующей температуры для черной дыры закрыло большой теоретический пробел в статистической механике черных дыр. Ранее он доказал важную теорему [45], согласно которой в классической общей теории относительности площадь горизонта будущего черной дыры никогда не может уменьшаться. Руководствуясь аналогией этого результата с вторым началом термодинамики, Бекенштейн [5] попытался установить, нельзя ли эту площадь рассматривать как меру энтропии. Подсчитывая информацию, которая теряется внешним миром, когда черная дыра захватывает материю или излучение, он нашел, что площадь действительно служит такой мерой и коэффициент пропорциональности между ней и термодинамической энтропией должен быть очень близок к единице в системе единиц, в которой $\hbar=c=G=k=1$.

Однако догадка Бекенштейна имела один явный недостаток. Классически черная дыра может быть в тепловом равновесии со своим окружением только в том случае, если температура этого окружения равна абсолютному нулю, а это означает, что сама дыра должна быть при нулевой температуре. Если энтропию S считать пропорциональной площади, то из термодинамической формулы $dE=TdS$ следует, что для конечного изменения энергии E (или массы M) черной дыры требуется бесконечное изменение площади. Это находится в прямом противоречии с элементарным соотношением между массой и площадью $A=16\pi M^2$ (для невращающейся черной дыры).

Если же черной дыре приписать температуру, определяемую формой (15), то упомянутое термодинамическое соотношение принимает простую форму $dM=1/4TdA$ и допускает отождествление [1])

$$S=\frac{1}{4}A. \qquad (16)$$

Открытие Хокинга не только устанавливает коэффициент пропорциональности между энтропией и площадью горизонта, но также показывает самым убедительным образом, что общая теория относительности должна сочетаться с квантовой теорией, если мы хотим обеспечить согласие со статистической механикой.

Следует особо подчеркнуть общность приведенных результатов: хотя температура (15) была первоначально выведена при исследовании *линейных* квантованных полей, распространяющихся в заданной геометрии черной дыры, она верна также и для взаимодействующих полей. Наиболее очевидно это следует из работ кембриджской школы (см. [48*] и статью Хокинга VII в настоящем сборнике). В них показано, что одно из важных свойств хронологических функций Грина для любой теории поля (с взаимодействием или без него) на фоне черной дыры может быть выведено просто аналитическим продолжением метрики к комплексным значениям времени t. Оказывается, что эти функции должны быть периодическими (для фермионов антипериодическими) по t с мнимым периодом i/T, если они «хорошо себя ведут» на горизонте и обладают нужными свойствами на бесконечности. А такая периодичность, как известно, является свойством тепловой функции Грина (см. [36]). Это можно усмотреть, например, в формуле (13), где период равен $2\pi i/a$.

3. РЕАКЦИЯ ИЗЛУЧЕНИЯ

Все фундаментальные поля, включая гравитацию, дают вклад в тепловое излучение, исходящее от черной дыры. Поэтому о самой черной дыре правильно говорить как о квантовом объекте. Квантованная черная дыра является, очевидно, динамическим объектом, который может обмениваться энергией и энтропией со своим окружением. Этот обмен всегда происходит с соблюдением законов статистической механики при условии, что они понимаются как законы для всей системы в целом, включающей черную дыру и ее окружение. Однако тот факт, что черная дыра может терять энтропию, означает, что в квантовой теории ее площадь может уменьшаться.

---

[1]) Соотношение (16) выполняется также для вращающихся заряженных черных дыр. Для них термодинамическая формула обобщается и принимает вид $dM=\frac{1}{4}TdA+\Omega dJ+\Phi dQ$, где Ω, J, Φ, Q — угловая скорость, момент импульса, электростатический потенциал и заряд соответственно. Если такие черные дыры образуются при коллапсе, то излучение на бесконечности не является чисто тепловым, но включает компоненту спонтанного излучения, возникающего в нейтральном случае из-за явления эргосферного сверхизлучения [12, 62, 73], а в заряженном случае еще и из-за парадокса Клейна.

Из известных выражений для площади и температуры легко вывести, что черная дыра, образовавшаяся при коллапсе, имеет светимость порядка M^{-2}. Если допустимо предположить, что светимость остается пропорциональной M^{-2} при вековых изменениях по M, то M постоянно убывает до тех пор, пока, в конце концов, черная дыра не исчезнет со взрывом или не будет стабилизирована каким-либо еще не известным нам квантовым эффектом. Время жизни для этого процесса порядка M^3.

Светимость дает информацию относительно тензора натяжений $\langle \mathbf{T}^{\mu\nu} \rangle$ только на бесконечности. Для того чтобы изучить детальное поведение геометрии во времени вблизи (и даже внутри) горизонта, мы должны определить $\langle \mathbf{T}^{\mu\nu} \rangle$ и в этой области. Для этого необходимо преодолеть ряд практических трудностей. 1) Точный вид базисных функций для геометрии Шварцшильда не известен; еще меньше мы знаем о них в случае геометрии Керра. 2) Для перенормировки $\langle \mathbf{T}^{\mu\nu} \rangle$ здесь недостаточно вычесть расходимости для пространства Минковского: появляются и иные расходимости, связанные с кривизной. 3) Какая бы схема перенормировки ни применялась, она должна быть ковариантной и применимой в общем случае; нельзя подгонять ее специально к геометрии черной дыры.

В последующих разделах будет вкратце изложен один из стандартных методов перенормировки $\langle \mathbf{T}^{\mu\nu} \rangle$. Но сначала необходимо установить более точно ту роль, которую, как это представляется, играет $\langle \mathbf{T}^{\mu\nu} \rangle$. Пусть $\boldsymbol{\psi}$ — коллективное обозначение всех форм вещества и (или) полей излучения, присутствие которых предполагается. В квантовой теории $\boldsymbol{\psi}$ есть квантовый оператор. Метрический тензор $\mathbf{g}_{\mu\nu}$ также является оператором. Допустим, что его можно разбить на (пока произвольный) классический фон $g_{\mu\nu}$ и операторный остаток $\boldsymbol{\phi}_{\mu\nu}$, причем и тот и другой рассматриваются как формальные тензоры. Общепринятое выражение для действия S совокупности вещества, полей излучения и гравитации содержит $g_{\mu\nu}$ и $\boldsymbol{\phi}_{\mu\nu}$ только в виде комбинации $g_{\mu\nu}+\boldsymbol{\phi}_{\mu\nu}$. Введем альтернативный функционал

$$\bar{S}[g, \boldsymbol{\phi}, \boldsymbol{\psi}] \overset{\text{def}}{=} S[g+\boldsymbol{\phi}, \boldsymbol{\psi}] - S[g, 0] - \int \frac{\delta S[g, 0]}{\delta g_{\mu\nu}} \boldsymbol{\phi}_{\mu\nu} d^n x, \qquad (17)$$

где (при обобщении на n-мерный случай) $d^n x$ обозначает $dx^0 dx^1 \ldots dx^{n-1}$. Если отстуствуют какие-либо внешние источники и, следовательно, $\boldsymbol{\psi}=0$ есть решение классических уравнений вещества и излучения и если выбор $g_{\mu\nu}$ таков, что удовлетворяются классические уравнения в пустом пространстве $\delta S[g, 0]/\delta g_{\mu\nu}=0$ [1]),

1) Часто бывает удобно изменить нулевые точки для полей вещества и излучения так, чтобы они тоже вносили вклад в фон. Если через ψ_B обозначен фон вещества и излучения, то такой фон должен удовлетворять уравнениям

$$\delta S[g, \psi_B]/\delta g_{\mu\nu}=0, \qquad \delta S[g, \psi_B]/\delta \psi_B=0.$$

Под ψ_B, соответствующими фермионным полям, следует понимать антикоммутирующие грассмановы числа.

то операторные полевые уравнения могут быть записаны в виде

$$\frac{\delta \bar{S}[g, \phi, \psi]}{\delta \psi} = 0, \tag{18}$$

$$F^{\mu\nu\sigma\tau}\phi_{\sigma\tau} = -\frac{1}{2}\mathbf{T}^{\mu\nu}, \tag{19}$$

где

$$F^{\mu\nu\sigma\tau}\phi_{\sigma\tau} \overset{\text{def}}{=} \int \frac{\delta^2 S[g, 0]}{\delta g_{\mu\nu}(x)\,\delta g_{\sigma\tau}(x')} \phi_{\sigma\tau}(x')\, dx', \tag{20}$$

$$\mathbf{T}^{\mu\nu} \overset{\text{def}}{=} 2\,\frac{\delta \bar{S}[g, \phi, \psi]}{\delta g_{\mu\nu}}, \tag{21}$$

$F^{\mu\nu\sigma\tau}$ — ковариантный самосопряженный дифференциальный оператор второго порядка, который зависит только от фоновой метрики. (Явный его вид нам не понадобится.) Вся нелинейность уравнений гравитационного поля сконцентрирована в правой части уравнения (19), которая выражает принцип обратной связи: полевой оператор $\phi_{\mu\nu}$ через $\mathbf{T}^{\mu\nu}$ служит отчасти источником самого себя. Формально можно рассматривать $\phi_{\mu\nu}$ и ψ на равном основании. Они взаимодействуют друг с другом и с самими собой при своем распространении в фоновой геометрии, описываемой метрикой $g_{\mu\nu}$. В физических условиях, в которых допустимо считать $\phi_{\mu\nu}$ и ψ «малыми», как $\bar{S}[g, \phi, \psi]$, так и $\mathbf{T}^{\mu\nu}$ могут быть представлены как функциональные степенные ряды по $\phi_{\mu\nu}$ и ψ, начинающиеся с квадратичных членов. Эти степенные ряды образуют основу для теории возмущений.

При выполнении уравнений (18) и (19) $\mathbf{T}^{\mu\nu}$ удовлетворяет условию

$$\mathbf{T}^{\mu\nu}{}_{;\nu} = 0, \tag{22}$$

где ковариантная производная берется относительно фоновой метрики. $\mathbf{T}^{\mu\nu}$ обладает всеми свойствами тензора натяжений и может рассматриваться в качестве такового [1]). Его (перенормированное) среднее значение и есть та величина, которая требуется в проблеме обратного воздействия. Она появляется в соотношении, которое получается из (19) взятием среднего значения:

$$F^{\mu\nu\sigma\tau}\langle\phi_{\sigma\tau}\rangle = -\frac{1}{2}\langle\mathbf{T}^{\mu\nu}\rangle. \tag{23}$$

Об этом уравнении следует сделать несколько замечаний. Во-первых, в последующем будет удобно всегда заменять обычные

---

[1]) В действительности при определении формулой (21) $\mathbf{T}^{\mu\nu}$ является тензорной *плотностью* веса 1. В данном обзоре всюду удобно брать тензор натяжений в форме плотности.

средние значения «швингеровскими» средними

$$\langle \mathbf{A} \rangle \overset{\text{def}}{=} \frac{\langle \text{out, vac} \,|\, T(\mathbf{A}) \,|\, \text{in, vac} \rangle}{\langle \text{out, vac} \,|\, \text{in, vac} \rangle}, \tag{24}$$

где T — оператор хронологического упорядочения. Если $\mathbf{A}$ — локальный оператор, так что символ T может быть опущен, и нет рождения частиц фоновым полем, так что вакуумные векторы состояния in и out тождественны, то швингеровское среднее совпадает с обычным вакуумным средним. В более общем случае эти два средних различаются, но обычно на конечную величину, даже если оператор $\mathbf{A}$ не перенормирован. Следовательно, при исследовании вопросов перенормировки мы можем ограничиться рассмотрением только швингеровского среднего. Мы пока отложим вопрос о том, как определить швингеровское среднее, когда состояния in и out не могут быть введены общепринятым способом из-за отсутствия времениподобных векторов Киллинга или когда они неоднозначны по другим причинам.

Во-вторых, следует заметить, что хотя операторы $\boldsymbol{\phi}_{\mu\nu}$ и $\mathbf{T}^{\mu\nu}$ можно считать преобразующимися как обычные тензоры при координатных преобразованиях фоновой метрики, они *не* инвариантны относительно калибровочных преобразований квантовой теории гравитации. При калибровочных преобразованиях фоновая метрика остается фиксированной, и когда квантовый метрический тензор подвергается калибровочному преобразованию [см. равенство (95)], вся «тяжесть» этого изменения «ложится» на $\boldsymbol{\phi}_{\mu\nu}$. Поэтому средние значения (23) неоднозначны. Позднее мы введем особый способ устранения этой неоднозначности [см. (127), (132), (133)], который состоит во взятии гауссова среднего по калибровкам специального вида. Если такая процедура усреднения проведена, в (23) [см. формулу (158)] должны появиться дополнительные члены, но для приближений, рассматриваемых в разделах, следующих непосредственно за этим разделом; уравнение (23) справедливо в приведенном здесь виде.

Стоит заметить, что эта процедура усреднения во многих отношениях аналогична методу пространственно-временного усреднения, введенному Айзексоном [52] в классической теории для получения корректно определенного тензора натяжений. Бор и Розенфельд [7] в свое время подчеркивали, что физический смысл имеют только пространственно-временные усреднения полевых величин. Усредненный тензор натяжений может «наблюдаться» только путем измерений в его окрестности пространственно-временных средних «локальной компоненты» тензора Римана, по существу через измерение левой части (23). (Тензор натяжений не измеряется детектором частиц!) Если для гравитационного поля провести анализ [19], подобный анализу Бора — Розенфельда, то окажется, что средние значения локального тензора Римана в квантовой области могут быть измерены (в принципе) с достаточной точностью, пока область

усреднения больше планковской длины. Поэтому уравнение (23) имеет настоящий операционный смысл.

Математически точное утверждение относительно этого смысла можно получить, вводя следующие определения:

$$\varphi_{\mu\nu} \overset{\text{def}}{=} g_{\mu\nu} + \langle \phi_{\mu\nu} \rangle = \langle g_{\mu\nu} \rangle, \qquad \psi \overset{\text{def}}{=} \langle \psi \rangle. \tag{25}$$

Ниже будет показано, что существует функционал $\Gamma[\varphi, \psi]$ от $\varphi_{\mu\nu}$ и ψ, такой, что (исправленным) уравнениям (23) и швингеровскому среднему от уравнения(18) может быть придана соответственно следующая форма:

$$\frac{\delta\Gamma}{\delta\varphi_{\mu\nu}} = 0, \tag{26}$$

$$\frac{\delta\Gamma}{\delta\psi} = 0. \tag{27}$$

Здесь Γ— так называемое *эффективное действие*. Оно описывает динамическое поведение когерентных гравитационных, электромагнитных и прочих полей большой амплитуды. При таком описании роль основной «классической» метрики играет $\varphi_{\mu\nu}$, а не $g_{\mu\nu}$.

Поскольку поведение $\varphi_{\mu\nu}$ определяется функционалом Γ, а не S, то в него включено и самодействие. Использование $\varphi_{\mu\nu}$ вместо $g_{\mu\nu}$ позволяет описать в квазиклассических выражениях *распадающуюся* черную дыру (по крайней мере до тех пор, пока мнимая часть $\varphi_{\mu\nu}$ остается малой). Поскольку значение $\varphi_{\mu\nu}$ определяется уравнением (23), роль $\langle \mathbf{T}^{\mu\nu} \rangle$ теперь очевидна.

Когда рождение частиц становится очень интенсивным, как, например, в конечной стадии распада черной дыры, мнимая часть $\varphi_{\mu\nu}$ уже не мала, и квазиклассическая картина пререстает быть верной. Вопрос о корректной интерпретации *комплексной* $\varphi_{\mu\nu}$ остается открытым. Пока не ясно, можно ли решить этот вопрос в рамках общей философии аналитического продолжения или же комплексные $\varphi_{\mu\nu}$ должны играть какую-то роль в теории инстантонов и в топологических проблемах. Следует отметить, однако, что Γ в любом случае имеет смысл в терминах амплитуд перехода. Например, полная S-матрица данной теории определяется «древесными амплитудами» функционала Γ, а они могут быть введены относительно любого асимптотически плоского фона [21].

4. ОДНОПЕТЛЕВОЕ ПРИБЛИЖЕНИЕ

В низшем приближении правая часть уравнения (23) может быть найдена путем подстановки в выражение для $\mathbf{T}^{\mu\nu}$ подходящих операторных решений уравнений $F^{\mu\nu\sigma\tau}\phi_{\sigma\tau}=0$ и $(\delta\bar{S}[g, 0, \psi]/\delta\psi)_{\text{лин}} = 0$, где «лин» означает «сохранить только часть, линейную

по ψ». В этом приближении поля $\phi_{\mu\nu}$ и ψ не взаимодействуют друг с другом (и с самими собой), а лишь линейно распространяются на фоне $g_{\mu\nu}$. Для самосогласованности (на этом уровне) достаточно сохранить в разложении $\mathbf{T}^{\mu\nu}$ в ряд только члены, квадратичные по $\phi_{\mu\nu}$ и ψ. Получаемое значение $\langle\mathbf{T}^{\mu\nu}\rangle$ известно как *однопетлевое приближение*. Именно это приближение имеют обычно в виду, когда рассматривают тензор натяжений. Оно лежит в основе известных результатов, упомянутых в предыдущих разделах. К сожалению, из-за того, что квантовая гравитация технически неперенормируема, удовлетворительный во всех отношениях способ придать смысл этой теории в более высоких порядках пока не известен.

В этом обзоре перенормировка $\langle\mathbf{T}^{\mu\nu}\rangle$ рассматривается только в однопетлевом приближении. Хотя при этом наше понимание квантовой гравитации будет неполным, следует заметить, что однопетлевое приближение в полевой теории эквивалентно ВКБ-приближению обычной квантовой механики. Поэтому есть основания ожидать, что, как и в обычной квантовой механике, оно приведет по меньшей мере к прояснению некоторых существенных сторон точной теории.

При исследовании однопетлевого приближения удобно ввести функционал $W[g]$ фоновой метрики, определяемый равенством

$$e^{iW} \overset{\text{def}}{=} \langle \text{out, vac} | \text{in, vac} \rangle. \tag{28}$$

Из уравнений (1), (21), (24) следует

$$\langle\mathbf{T}^{\mu\nu}\rangle = 2\frac{\delta W}{\delta g_{\mu\nu}}. \tag{29}$$

Перенормировка $\langle\mathbf{T}^{\mu\nu}\rangle$, очевидно, может быть достигнута перенормировкой W. В действительности только такая процедура гарантирует от ошибок. Вспоминая, что фоновая метрика $g_{\mu\nu}$ удовлетворяет уравнениям поля в пустом пространстве, мы видим, что из (23) и (29) следует уравнение

$$\frac{\delta S[g,0]}{\delta g_{\mu\nu}} + F^{\mu\nu\sigma\tau}\langle\phi_{\sigma\tau}\rangle + \frac{\delta W[g]}{\delta g_{\mu\nu}} = 0. \tag{30}$$

В однопетлевом приближении оно эквивалентно уравнению

$$\frac{\delta S[\varphi,0]}{\delta\varphi_{\mu\nu}} + \frac{\delta W[\varphi]}{\delta\varphi_{\mu\nu}} = 0, \tag{31}$$

которое позволяет в данном порядке отождествить сумму $S[\varphi, 0] + W[\varphi]$ с частью эффективного действия, не зависящей от ψ. Поскольку именно эффективное действие описывает *реальную* физику данной теории, то отсюда следует, что ни S, ни W по отдельности не имеют физического смысла, а имеет смысл только их сумма $S+W$. Если W содержит члены, подобные членам, имеющимся в S, то только суммы соответствующих коэффициентов могут быть определены экспериментально как наблюдаемые «константы взаимодей-

ствия». Если какие-либо члены в W имеют расходящиеся коэффициенты, они могут быть скомпенсированы «контрчленами» в S. Вообще говоря, могут быть также и конечные члены, подлежащие компенсации.

В однопетлевом приближении расходимости операторного тензора натяжений $\mathbf{T}^{\mu\nu}$ такие же, как у его швингеровского среднего, и потому являются c-числами. Перенормированный тензор натяжений определяется как

$$\mathbf{T}^{\mu\nu}_{\text{ren}} \stackrel{\text{def}}{=} \mathbf{T}^{\mu\nu} - 2\,\frac{\delta \Delta W}{\delta g_{\mu\nu}}, \tag{32}$$

где ΔW — максимальная часть W, которая может быть скомпенсирована контрчленами в S, удовлетворяющими следующим условиям: 1) используются только такие типы контрчленов, которые нужны для того, чтобы $\mathbf{T}^{\mu\nu}_{\text{ren}}$ был свободен от расходимостей; 2) $\langle \mathbf{T}^{\mu\nu}_{\text{ren}} \rangle$ должен обращаться в нуль в плоском пустом пространстве-времени; 3) выражение для ΔW должно быть координатно-инвариантным; 4) производная $\delta \Delta W / \delta g_{\mu\nu}$ должна зависеть от фоновой геометрии только локально. Координатная инвариантность ΔW гарантирует, что условие (22) останется справедливым после проведения перенормировки. Это одно из наиболее важных оснований для того, чтобы перенормировать сначала W, а затем получить $\mathbf{T}^{\mu\nu}_{\text{ren}}$ как производную величину. Если иметь дело непосредственно с $\mathbf{T}^{\mu\nu}$, всегда есть опасность, что условие (22) будет нарушено.

4.1. ФОРМАЛЬНЫЕ СООТНОШЕНИЯ; РОЛЬ ФЕЙНМАНОВСКОГО ПРОПАГАТОРА

Для того чтобы вычесть из W координатно-инвариантную расходящуюся часть, нужно сначала однозначным образом выделить ее. Это требует введения общековариантной схемы регуляризации. Чтобы построить такого рода схему, мы должны начать с некоторых формальных соотношений. Допустим на некоторое время, что мы интересуемся скалярным полем ϕ. Тогда $\langle \mathbf{T}^{\mu\nu} \rangle$ формально получается сложением соответствующих производных швингеровского среднего $\langle \phi(x)\phi(x') \rangle$ и переходом к пределу $x' \to x$. Если потребовать, чтобы x' стремилось к x по пространственноподобному направлению, это среднее может быть связано с фейнмановским пропагатором $G(x, x')$ [1]:

$$\langle \phi(x)\,\phi(x') \rangle = -iG(x, x'). \tag{33}$$

Фейнмановский пропагатор есть функция Грина, удовлетворяющая уравнению

$$FG(x, x') = -\delta(x, x'), \tag{34}$$

[1]) В общем случае в левой части равенства (33) должно стоять хронологическое произведение.

где F — дифференциальный оператор, присутствующий в уравнении поля $F\phi=0$. Решающее значение для теории имеют граничные условия, определяющие функцию Грина. Когда фоновая геометрия допускает глобальный времениподобный вектор Киллинга, относительно которого определяются вакуумные состояния (в этом случае |in, vac>=|out, vac>), $G(x, x')$ можно представить как сумму соответствующих базисных функций. Если $G(x, x')$ рассматривается как функция x при фиксированном x', то ее можно разложить на чисто положительно-частотные базисные функции, когда x лежит в будущем точки x', и на чисто отрицательно-частотные базисные функции, когда x лежит в прошлом x'[1]). То же верно, если $G(x, x')$ рассматривается как функция x' при фиксированном x, поскольку она симметрична по своим аргументам.

Можно показать, что эти свойства получаются из следующего формального выбора решения уравнения (34):

$$G = -\frac{1}{F+i0}. \tag{35}$$

Здесь не обозначены аргументы x и x', а символ $i0$ означает добавление единичного оператора с малым мнимым положительным множителем и переход к нулевому пределу по этому множителю. Альтернативный способ получения той же функции Грина состоит в повороте временнóй координаты x^0 в комплексной плоскости на 90° по часовой стрелке; при этом оператор F, который в скалярном случае имеет вид $g^{1/2}(\Box-\xi R-m^2)$[2]), становится отрицательно определенным. Такой оператор обладает единственной функцией Грина, обращающейся в нуль, когда x и x' разделены бесконечным интервалом. Фейнмановский пропагатор получается аналитическим продолжением этой функции Грина обратно к физическим значениям времени.

Когда глобального времениподобного векторного поля нет нигде, кроме областей «in» и «out», фейнмановский пропагатор определяется требованием разложимости (по каждому из двух аргументов) по отрицательно-частотным базисным функциям в области «in» и по положительно-частотным базисным функциями в области «out». Глобально-гиперболическое пространство-время, обладающее областями «in» и «out», может быть получено из пространства-времени, обладающего глобальным времениподобным вектором Киллинга,

---

[1]) Когда x и x' разделены пространственноподобным интервалом, возможно любое из этих двух разложений. Если имеются моды нулевой частоты (например, когда вселенная компактна, а поле безмассовое), настоящего фейнмановского пропагатора не существует, и пространство состояний не может быть построено как строгое пространство Фока. Методы, необходимые для такого случая, в данном обзоре не рассматриваются.

[2]) Здесь $g \overset{\text{def}}{=} -\det(g_{\mu\nu})$, $\Box$ — оператор Лапласа — Бельтрами, ξ — постоянная, R — скалярная кривизна, m — масса скалярного поля, а сигнатура метрики есть $-++\ldots$.

интегрированием бесконечной последовательности инфинитезимальных вариаций метрики $\delta g_{\mu\nu}$. Требуемое граничное условие выполняется, если при каждой из этих вариаций G получает изменение

$$\delta G = G \delta F G, \tag{36}$$

где δF — соответствующее изменение F. Поскольку F, а следовательно, и δF самосопряжены, симметрия $G(x, x')$ по аргументам остается неизменной. Что еще более важно, при этом сохраняется представление (35).

Естественно задаться вопросом: существует ли функция Грина, удовлетворяющая (35) и (36) и обладающая нужным свойством симметрии, в тех случаях, когда вообще нет областей «in» и «out»? Общего ответа на этот вопрос пока не найдено, но известно, что такие функции Грина существуют в тех случаях, когда можно воспользоваться техникой глобального аналитического продолжения и (или) конформных преобразований, *включая* случаи, в которых имеются геометрические сингулярности. Например, среднее хронологического произведения $iT(\phi(x)\phi(x'))$ по естественному тепловому состоянию, связанному с черной дырой, имеет представление (35). В этом случае данная функция единственна. Следует ли вообще единственность из существования — это вопрос, который пока полностью не решен [1]).

Представление (35) может быть записано в виде

$$g^{1/4} G g^{1/4} = i \int_0^\infty \exp\left(i g^{-1/4} F g^{-1/4} s\right) ds, \tag{37}$$

где множители $g^{-1/4}$ введены для того, чтобы обеспечить общековариантность подынтегрального выражения. Вопрос о существовании G может быть заменен вопросом о существовании интеграла (37).

Восстанавливая явную запись зависимости от аргументов x и x', получаем эквивалентное (37) выражение

$$G(x, x') = i \int_0^\infty g^{-1/4}(x) K(x, x', s) g^{-1/4}(x') ds, \tag{38}$$

где функция K удовлетворяет дифференциальному уравнению

$$\frac{\partial}{i\partial s} K(x, x', s) = g^{-1/4} F g^{-1/4} K(x, x', s) \tag{39}$$

и «начальному» условию

$$K(x, x', 0) = \delta(x, x'). \tag{40}$$

Когда пространство-время неполное, функция K должна также удовлетворять определенным условиям на границе. В проблеме Кази-

[1]) Из существования действительно следует единственность, когда пространственно-временнóе многообразие является полным, а от соответствующих базисных функций требуется, чтобы они были ограниченны.

мира эти условия элементарны, но если неполнота обусловлена геометрическими сингулярностями, они могут включать аналитическое продолжение и (или) конформное преобразование метрики.

Теперь вклад скалярного поля в классическое действие может быть записан в виде

$$S_{\text{ск}}[\phi]=\frac{1}{2}\int \phi F\phi\, d^n x. \tag{41}$$

Поскольку $T^{\mu\nu}=2\delta S_{\text{ск}}/\delta g_{\mu\nu}$, из (29) следует, что при вариации метрики $\delta g_{\mu\nu}$ и соответствующем изменении δF оператора F функционал W получает изменение

$$\delta W=\frac{1}{2}\int \langle\phi\delta F\phi\rangle\, d^n x=-\frac{i}{2}\,\mathrm{tr}(G\delta F), \tag{42}$$

где при переходе к этой окончательной форме использовано равенство (33), а интегрирование по пространству-времени заменено символом следа. Оператор δF является локальным, поэтому взятие следа связано с пределами совпадения (по пространственноподобным направлениям) не выписанных явно аргументов x и x'. Поскольку $F\phi=0$, равенство (42) можно также записать в виде

$$\delta W=\frac{1}{2}\int\langle g^{1/4}\phi\delta\,(g^{-1/4}Fg^{-1/4})\,g^{1/4}\phi\rangle d^n x=$$

$$=-\frac{i}{2}\,\mathrm{tr}\,[g^{1/4}Gg^{1/4}\delta\,(g^{-1/4}Fg^{-1/4})], \tag{42'}$$

который позволяет использовать представление (37):

$$\delta W=\delta\left[\frac{1}{2}\,\mathrm{tr}\int_0^\infty \frac{1}{is}\exp\,(ig^{-1/4}Fg^{-1/4}s)\,ds\right]. \tag{43}$$

Это вариационное уравнение непосредствено интегрируется; получаем следующее формальное выражение через функцию K:

$$W=\frac{1}{2}\int_0^\infty \frac{1}{is}K\,(s)\,ds+\mathrm{const}, \tag{44}$$

$$K\,(s)\overset{\mathrm{def}}{=}\int K\,(x,\ x,\ s)\,d^n x. \tag{45}$$

4.2. ПРЕДСТАВЛЕНИЯ ФУНКЦИИ $K(x,x',s)$

Когда x и x' близки, удобно ввести представление функции K, которое подсказывается ее формой в плоском пространстве-времени:

$$K\,(x,\ x',\ s)=\frac{iD^{1/2}\,(x,\ x')}{(4\pi is)^{n/2}}\,e^{(i/2s)\,\sigma\,(x,\ x')-im^2 s}\Lambda\,(x,\ x',\ s). \tag{46}$$

Здесь n — размерность пространства-времени, $\sigma\,(x,\,x')$ — половина квадрата геодезического расстояния между x и x', $D\,(x,x')$—

определитель $n \times n$:

$$D(x, x') \overset{\text{def}}{=} -\det\left[-\frac{\partial^2\sigma(x, x')}{\partial x^\mu \partial x'^\nu}\right], \tag{47}$$

а $\Lambda(s, s', s)$ удовлетворяет граничному условию

$$\Lambda(x, x, 0) = 1. \tag{48}$$

Подставляя (46) и (39), вспоминая, что $F = g^{1/2}(\Box - \xi R - m^2)$, и используя следующие уравнения, которым удовлетворяют σ и D (см. [21]):

$$\sigma_{;\mu}\sigma_{;}{}^{\mu} = \sigma_{;\mu'}\sigma_{;}{}^{\mu'} = 2\sigma, \quad (D^{1/2}\sigma_{;}{}^{\mu})_{;\mu} = n, \tag{49}$$

находим

$$\frac{\partial}{i\partial s}\Lambda + \frac{1}{is}\sigma_{;}{}^{\mu}\Lambda_{;\mu} = D^{-1/2}(\Box - \xi R)(D^{1/2}\Lambda). \tag{50}$$

Когда пространство-время неполно, уравнения (50) вместе с граничным условием (48) недостаточно для определения Λ. Однако их достаточно для определения асимптотического разложения Λ, справедливого при малых s и x, близких к x':

$$\Lambda(x, x', s) \sim \sum_{r=0}^{\infty} a_r(x, x')(is)^r. \tag{51}$$

Коэффициенты $a_r(x, x')$ определяются дифференциальными рекуррентными соотношениями

$$a_0(x, x') = 1,$$

$$\sigma_{;}{}^{\mu}a_{r;\mu} + ra_r = D^{-1/2}(\Box - \xi R)(D^{1/2}a_{r-1}), \quad r = 1, 2, 3, \ldots, \tag{52}$$

которые следуют из (48) и (50). Повторно дифференцируя уравнение (49) и используя тот факт, что

$$\sigma_{;\mu} \underset{x' \to x}{\longrightarrow} 0, \quad \sigma_{;\mu\nu'} \underset{x' \to x}{\longrightarrow} -g_{\mu\nu},$$

находим, что из этих рекуррентных соотношений следуют пределы совпадения [1])

$$a_1(x, x) = \left(\frac{1}{6} - \xi\right)R, \tag{53}$$

$$a_2(x, x) = \frac{1}{2}\left(\frac{1}{6} - \xi\right)^2 R^2 + \frac{1}{180}(-R_{\mu\nu}R^{\mu\nu} + R_{\mu\nu\sigma\tau}R^{\mu\nu\sigma\tau}) + \frac{1}{6}\left(\frac{1}{5} - \xi\right)\Box R. \tag{54}$$

Из уравнений (45), (46) и (51) получаем

$$K(s) \sim \frac{i}{(4\pi i s)^{n/2}} e^{-im^2 s} \sum_{r=0}^{\infty} A_r (is)^r, \tag{55}$$

[1]) Выражение для $a_3(x,x)$ см. в работе [42].

где A_r — формальные интегралы:

$$A_r \overset{\text{def}}{=} \int g^{1/2} a_r(x, x)\, d^n x. \qquad (56)$$

Из этого выражения видно, что интеграл в (44) в нижнем пределе расходится как

$$\frac{1}{2}(4\pi)^{-n/2}\left[\frac{2}{n}A_0\,(is)^{-n/2}+\frac{2}{n-2}(A_1-m^2A_0)\,(is)^{-(n/2)+1}+\ldots\right].$$

Если пространство-время имеет несингулярный неизотропный край, представление (55) не совсем верно: иногда $\Lambda(x, x, s)$ содержит члены вида $\exp[if(x)/s]$, где при x, лежащем близко к краю, функция $f(x)$ пропорциональна квадрату расстояния от края. Эти члены обладают существенными особенностями при $s=0$, но не появляются в асимптотическом разложении (51) [1]). Однако, будучи подставлены в интеграл (45), они дают два типа дополнительных членов к сумме в (55): члены, содержащие полуцелые степени is, и члены, содержащие целые степени, подобные тем, которые уже имеются в сумме. Члены первого типа возникают из-за эффекта «перенаселения» вблизи края, который снижает эффективную размерность пространства-времени на единицу. Члены второго типа включают внешнюю кривизну (вторую фундаментальную форму) края, и их легче всего получить «удвоением» многообразия, т. е. пополняя его присоединением к нему вдоль края его метрической копии (см. [61]). За исключением эффекта краевого перенаселения, удвоенное многообразие обладает всеми основными свойствами исходного многообразия, но поскольку оно полное, выражение (55) может быть применено к нему в том виде, в котором оно приведено. Компоненты связности (в некоторой естественной системе координат) имеют разрывы по «стыку» вдоль исходного края, а тензор Римана ведет себя здесь как дельта-функция. Именно эта дельта-функция порождает поправки к (55) от внешней кривизны.

Мы видим, что искривленные края приводят к определенным нереалистическим и нефизическим эффектам уже в плоском пространстве-времени, например к бесконечным значениям плотности энергии у края. В общем случае они переплетаются со стандартными расходимостями и усложняют проблему перенормировки совершенно безнадежным образом. В дальнейшем мы исключим из рассмотрения многообразия с краем, если только этот край не возникает естественным образом при аналитическом продолжении или конформном преобразовании из многообразий без края. Тогда выражение (55) не нуждается в каких-либо изменениях.

---

[1]) Часто имеются также члены, содержащие множители вида $\exp(i\cdot\mathrm{const}/s)$ Такие члены могут появляться даже в том случае, когда многообразие не имеет края (например, если гиперповерхности Коши компактны), и также выпадают из асимптотического разложения, но они не вносят вклада ни в выражение (55), ни в расходимости интеграла (44).

Другое полезное представление функции $K(x, x', s)$ можно получить, вводя полный набор собственных функций $v(\alpha, x)$ оператора $-\Box+\xi R$:

$$(-\Box+\xi R)\, v(\alpha, x)=\lambda(\alpha)\, v(\alpha, x). \tag{57}$$

Без потери общности можно считать, что $v(\alpha, x)$ действительны и ортонормированны:

$$\begin{gathered}\int g^{1/2}(x)\, v(\alpha, x)\, v(\alpha', x)\, d^n x=\delta_{\alpha\alpha'},\\ \sum_\alpha g^{1/4}(x)\, v(\alpha, x)\, v(\alpha, x')\, g^{1/4}(x')=\delta(x, x').\end{gathered} \tag{58}$$

Если пространство-время полное (и глобально-гиперболическое), то функция v должна быть ограниченной; если же оно не полное, то функции должны удовлетворять определенным краевым условиям. Здесь α обозначает полный набор нумерующих величин, а символы $\delta_{\alpha\alpha'}$ и Σ следует понимать в том смысле, что в случае, когда величины α непрерывны, они включают дельта-функции и интегралы соответственно. С помощью функций v можно написать

$$K(x, x', s)=\sum_\alpha g^{1/4}(x)\, v(\alpha, x)\, v(\alpha, x')\, g^{1/4}(x')\, e^{-i[\lambda(\alpha)+m^2]s}, \tag{59}$$

$$K(s)=\sum_\alpha e^{-i[\lambda(\alpha)+m^2]s}. \tag{60}$$

Фейнмановский пропагатор может быть построен подобным же образом:

$$G(x, x')=\sum_\alpha \frac{v(\alpha, x)\, v(\alpha, x')}{\lambda(\alpha)+m^2-i0}. \tag{61}$$

4.3. ОБОБЩЕННАЯ ДЗЕТА-ФУНКЦИЯ

Свойства эфира (по крайней мере те свойства, за которые ответственно скалярное поле) полностью определяются оператором F и топологией пространства-времени. При анализе взаимосвязи этих двух факторов (например, при изучении распределения собственных значений F) математики находят удобным ввести обобщение дзета-функции Римана, определяемой следующим образом:

$$\zeta(x, z)\overset{\text{def}}{=}(g^{1/4}Gg^{1/4})^z(x, x)=\sum_\alpha \frac{g^{1/2}(x)\,[v(\alpha, x)]^2}{[\lambda(\alpha)+m^2-i0]^z}, \tag{62}$$

$$\zeta(z)\overset{\text{def}}{=}\operatorname{tr}(g^{1/4}Gg^{1/4})^z=\int \zeta(x, z)\, d^n x=\sum_\alpha \frac{1}{[\lambda(\alpha)+m^2-i0]^z}. \tag{63}$$

Ряд в (62) должен вычисляться в тех областях комплексной плоскости z, где он сходится, тогда в остальных областях $\zeta(x, z)$ опреде-

ляется аналитическим продолжением. Сама $\zeta(z)$ должна рассматриваться как формальный интеграл по пространству-времени [1]).

Определения (62) и (63) строго справедливы только при использовании безразмерных единиц (т. е. единиц, в которых $\hbar=c=G=1$), так что $g^{1/4}Gg^{1/4}$, $\lambda(\alpha)$, m^2 и т. п. безразмерны. В противном случае в эти определения необходимо ввести масштабный параметр. Мы опустим здесь этот параметр, а позднее кратко отметим, какие необходимы изменения (если они вообще необходимы), когда он вводится. Как легко видеть, дзета-функции связаны с функцией K:

$$\zeta(x, z)=\frac{1}{\Gamma(z)}\int_0^\infty (is)^{z-1} K(x, x, s)\, ids=$$

$$=\frac{i}{(4\pi)^{n/2}}\frac{g^{1/2}}{\Gamma(z)}\int_0^\infty (is)^{z-1-n/2}e^{-im^2 s}\Lambda(x, x, s)\, ids, \qquad (64)$$

$$\zeta(z)=\frac{1}{\Gamma(z)}\int_0^\infty (is)^{z-1}K(s)\, ids=\frac{i}{(4\pi)^{n/2}}\frac{1}{\Gamma(z)}\int_0^\infty (is)^{z-1-n/2}e^{-im^2 s}\Lambda(s)\, ids, \qquad (65)$$

$$\Lambda(s)\overset{\text{def}}{=}\int g^{1/2}(x)\Lambda(x, x, s)\, d^n x\sim\sum_{r=0}^{\infty} A_r (is)^r. \qquad (66)$$

Эти соотношения доставляют информацию об аналитической структуре $\zeta(x, z)$. Интегрирование по частям в сочетании с аналитическим продолжением из области $\operatorname{Re} z>n/2$ дает

$$\zeta(x, z)=\begin{cases} \dfrac{i}{4\pi}\dfrac{g^{1/2}}{(z-1)\Gamma(z+1)}\displaystyle\int_0^\infty (is)^z\left(\frac{\partial}{i\partial s}\right)^2\left[e^{-im^2 s}\Lambda(x,x,s)\right] ids\; (n=2), & (67\text{а}) \\[2ex] \dfrac{i}{(4\pi)^{3/2}}\dfrac{g^{1/2}}{\left(z-\frac{1}{2}\right)\left(z-\frac{3}{2}\right)\Gamma(z)}\displaystyle\int_0^\infty (is)^{z-1/2}\left(\frac{\partial}{i\partial s}\right)^2\left[e^{-im^2 s}\Lambda(x, x, s)\right]\times & \\ \qquad\times\, ids\; (n=3), & (67\text{б}) \\[2ex] -\dfrac{i}{(4\pi)^2}\dfrac{g^{1/2}}{(z-1)(z-2)\Gamma(z+1)}\displaystyle\int_0^\infty (is)^z\left(\frac{\partial}{i\partial s}\right)^3\left[e^{-im^2 s}\Lambda(x, x, s)\right]\times & \\ \qquad\times\, ids\; (n=4) & (67\text{в}) \end{cases}$$

и т. д., откуда можно заключить, что, когда n четное, $\zeta(x, z)$ имеет простые полюсы в точках $z=1, 2, \ldots, n/2$, а когда n нечетное,— в точках $z=n/2-r$, $r=1,2, \ldots$; никаких других сингулярностей

---

[1]) На компактных многообразиях (а также в некоторых других случаях) этот интеграл действительно существует.

нет. Отметим, что

$$\zeta(x, 0)=\begin{cases} 0 \ (n \text{ нечетное}), \\ \frac{i}{(4\pi)^{n/2}} \frac{g^{1/2}}{(n/2)!} \left\{ \left(\frac{\partial}{i\partial s} \right)^{n/2} \left[e^{-im^2 s} \Lambda(x, x, s) \right] \right\}_{s=0} \ (n \text{ четное}) \end{cases} \tag{68}$$

Представляет интерес исследовать последнее соотношение в частном случае $m=0$ и $\xi=(n-2)/4(n-1)$. Уравнения скалярного поля тогда конформно-инвариантны, и при изменении метрического тензора $\delta g_{\mu\nu}=g_{\mu\nu}\delta\lambda$, как легко показать,

$$\delta(g^{-1/4} F g^{-1/4}) = -\frac{1}{2}\{g^{-1/4} F g^{-1/4}, \ \delta\lambda\},$$

откуда

$$\delta(g^{1/4} G g^{1/4}) = \frac{1}{2}\{g^{1/4} G g^{1/4}, \ \delta\lambda\}, \ \delta\zeta(z) = z \operatorname{tr}[(g^{1/4} G g^{1/4})^z \delta\lambda]. \tag{69}$$

В частности, $\delta\zeta(0)=0$. Но из (68) получаем

$$\zeta(0)=\left\{\begin{matrix} 0 \ (n \text{ нечетное}), \\ \frac{i}{(4\pi)^{n/2}} A_{n/2} \ (n \text{ четное}) \end{matrix}\right\} \left(m=0, \ \xi=\frac{1}{4}\frac{n-2}{n-1} \right), \tag{70}$$

откуда следует, что $A_{n/2}$ (n четное) конформно-инвариантно [1]).

4.4. РЕГУЛЯРИЗАЦИЯ И ПЕРЕНОРМИРОВКА

В дополнение к выражению (44) вариационное уравнение (42′) имеет также следующий формальный интеграл:

$$W = -\frac{i}{2} \operatorname{tr} \ln (g^{1/4} G g^{1/4}) + \text{const} = \tag{71а}$$

$$= \frac{i}{2} \sum_{\alpha} \ln[\lambda(\alpha) + m^2 - i0] + \text{const}. \tag{71б}$$

Сравнивая (63) и (71б), получаем еще одно выражение для W:

$$W = -\frac{i}{2} \zeta'(0) + \text{const}. \tag{72}$$

Между выражениями (44) и (72) есть важное различие. В то время как интеграл в (44) расходится в нижнем пределе, выражение (72) *конечно* [2]), в чем можно убедиться, дифференцируя (65) и аналитичееки продолжая из области $\operatorname{Re} z > n/2$ к $z=0$. Отбрасывая в

---

[1]) Автор признателен Дж. Даукеру за это доказательство.

[2]) Точнее говоря, оно представляет собой формальный интеграл по пространству-времени от конечной геометрической величины.

(72) постоянную интегрирования, находим

$$W=\begin{cases}\frac{1}{8\pi}\left\{(\gamma+1)(A_1-m^2A_0)-\int\limits_0^\infty \ln(is)\left(\frac{\partial}{i\partial s}\right)^2[e^{-im^2s}\Lambda(s)]\,ids\right\} & (n=2), \quad (73а)\\ \frac{1}{12\pi^{3/2}}\int\limits_0^\infty (is)^{-1/2}\left(\frac{\partial}{i\partial s}\right)^2[e^{-im^2s}\Lambda(s)]\,ids \quad (n=3), & (73б)\\ \frac{1}{32\pi^2}\left\{\left(\gamma+\frac{3}{2}\right)\left(A_2-m^2A_1+\frac{1}{2}m^4A_0\right)-\right. & \\ \qquad\left.-\frac{1}{2}\int\limits_0^\infty \ln(is)\left(\frac{\partial}{i\partial s}\right)^3[e^{im^2s}\Lambda(s)]\,ids\right\}(n=4) & (73в)\end{cases}$$

и т. д.

Этот метод получения конечного значения для W использовался Даукером и Критчли [27] и Хокингом [48] и известен как *регуляризация через дзета-функцию*. Есть другой распространенный метод, известный как *размерная регуляризация*, который приводит к весьма похожим результатам. В этом методе выражение (46) подставляется в (45) и (44), а W определяется аналитическим продолжением по размерности пространства-времени n. Когда физическая размерность нечетна, метод, использующий дзета-функцию, и размерный метод дают тождественные результаты. При четной физической размерности результаты по размерному методу отличаются от выражений (73) в членах, содержащих A_0, A_1, A_2 и т. д. Для того чтобы получить результаты размерной регуляризации, нужно в (73) множители, содержащие постоянную Эйлера γ, заменить определенными функциями n, каждая из которых имеет простой полюс при соответствующей физической размерности.

Явно выделенные величины A принято рассматривать как перенормировочные члены, которые должны быть поглощены контрчленами в классическом действии S. Поэтому обе регуляризационные схемы приводят к одним и тем же *физическим* результатам. Отметим, что при $n=4$ в классическое действие должен быть включен контрчлен, пропорциональный A_2. Такой контрчлен требует введения новой *неклассической* константы взаимодействия. О некоторой неоднозначности, которая возникает из-за присутствия такого члена, будет сказано ниже.

При простом отбрасывании в (73) перенормировочных членов мы не получаем полностью перенормированного W. Нужно еще исключить некоторые другие члены. Рассмотрим сначала случай $m\neq 0$. Интегрированием по частям можно непосредственно убедить-

ся, что

$$-\frac{1}{8\pi}\int_0^\infty \ln(is)\left(\frac{\partial}{i\partial s}\right)^2[e^{-im^2s}\Lambda(s)]\,ids =$$
$$= W_{\text{ren}} - \frac{1}{8\pi}(\gamma+\ln m^2)(A_1 - m^2A_0) - \frac{1}{8\pi}A_1 \quad (n=2), \tag{74}$$

$$\frac{1}{12\pi^{3/2}}\int_0^\infty (is)^{-1/2}\left(\frac{\partial}{i\partial s}\right)^2[e^{-im^2s}\Lambda(s)]\,ids =$$
$$= W_{\text{ren}} - \frac{m}{12\pi}\left(\frac{3}{2}A_1 - m^2A_0\right) \quad (n=3), \tag{75}$$

$$-\frac{1}{64\pi^2}\int_0^\infty \ln(is)\left(\frac{\partial}{i\partial s}\right)^3[e^{-im^2s}\Lambda(s)]\,ids =$$
$$= W_{\text{ren}} - \frac{1}{32\pi^2}(\gamma+\ln m^2)\left(A_2 - m^2A_1 + \frac{1}{2}m^2A_0\right) -$$
$$-\frac{1}{32\pi^2}\left(\frac{3}{2}A_2 - \frac{1}{2}m^2A_1\right) \quad (n=4), \tag{76}$$

где

$$W_{\text{ren}} = \begin{cases} \dfrac{1}{18\pi}\displaystyle\int_0^\infty \frac{1}{is^2}e^{-im^2s}[\Lambda(s) - A_0 - A_1(is)]\,ids \quad (n=2), & (77\text{а}) \\ \dfrac{1}{16\pi^{3/2}}\displaystyle\int_0^\infty \frac{1}{(is)^{5/2}}e^{-im^2s}[\Lambda(s) - A_0 - A_1(is)]\,ids \quad (n=3), & (77\text{б}) \\ \dfrac{1}{32\pi^2}\displaystyle\int_0^\infty \frac{1}{(is)^3}e^{-im^2s}[\Lambda(s) - A_0 - A_1(is) - A_2(is)^2]\,ids \quad (n=4). & (77\text{в}) \end{cases}$$

W_{ren} есть полностью перенормированное W. Отметим, что его можно было бы получить непосредственно из выражения (44), просто вычитая из $K(s)$ в каждом случае столько первых членов асимптотического разложения, сколько нужно для того, чтобы интеграл стал сходящимся. Эти вычитаемые члены можно считать соответствующими ΔW в формуле (32).

Этот простой вычитательный алгоритм применим также при $m=0$ и нечетном n. Но он не всегда применим при $m=0$ и четном n; при этом приходится исследовать каждый случай отдельно. Трудность здесь в том, что теперь в подынтегральных выражениях в (77) уже нет множителя e^{-im^2s}, поэтому нечем контролировать сходимость интеграла при верхнем пределе. Эта проблема возникает не из-за функции $\Lambda(s)$, поведение которой при верхнем пределе пред-

полагается хорошим, а из-за вычитаемого члена более высокой степени по is, а именно из-за члена в $A_{n/2}$.

Рассмотрим сначала случай $n=2$. Тогда по формуле Гаусса — Бонне A_1 пропорционально характеристике Эйлера — Пуанкаре данного пространства-времени. Поэтому оно не просто конформно-инвариантно, но также и *метрически-инвариантно* и не дает вклада в перенормированный тензор натяжений

$$\langle \mathbf{T}^{\mu\nu}_{\text{ren}} \rangle = 2\delta W_{\text{ren}} / \delta g_{\mu\nu}. \tag{78}$$

Это означает, что (77а) можно формально использовать в уже выписанном виде; при этом получим

$$\langle \mathbf{T}^{\mu\nu}_{\text{ren}}(x) \rangle = \frac{1}{8\pi} \int_0^\infty \frac{1}{(is)^2} [T^{\mu\nu}(x, s) - g^{1/2}(x) g^{\mu\nu}(x)] i ds \quad (m=0, n=2), \tag{79}$$

$$T^{\mu\nu}(x, s) \overset{\text{def}}{=} 2 \frac{\delta \Lambda(s)}{\delta g_{\mu\nu}(x)}. \tag{80}$$

Величина в квадратных скобках в формуле (79) порядка $(is)^2$ при $s \to 0$, и, следовательно, интеграл сходится.

Случай $n=4$ значительно сложнее. Положим $\xi = {}^1/_4 (n-2)/(n-1) = {}^1/_6$, так что уравнение поля будет конформно-инвариантным. Тогда $a_1(x, x) = 0$ (см. (53)), и член с A_1 выпадает из (77в). Член с A_2, напротив, не обращается в нуль. Более того, хотя интеграл A_2 и конформно-инвариантен, он все же не является метрически-инвариантным. Его функциональная производная по $g_{\mu\nu}$ не обращается в нуль, за исключением частных случаев. Наиболлее важными частными случаями являются те, в которых пространство-время либо конформно-плоское (тензор Вейля равен нулю), либо риччи-плоское (тензор Риччи равен нулю). В этих случаях выражением (77в) можно формально пользоваться в том виде, как оно написано; при этом получим

$$\langle \mathbf{T}^{\mu\nu}_{\text{ren}}(x) \rangle = \frac{1}{32\pi^2} \int_0^\infty \frac{1}{(is)^3} [T^{\mu\nu}(x, s) - g^{1/2}(x) g^{\mu\nu}(x)] i ds \tag{81}$$

($m=0$, $n=4$, $\xi = {}^1/_6$, пространство-время конформно-плоское или риччи-плоское). Величина в квадратных скобках теперь порядка (is^3) при $s \to 0$.

Если рассматривается конформно-неинвариантное уравнение или если пространство-время не является ни конформно-плоским, ни риччи-плоским, то выражением (77в) нельзя пользоваться даже формально. Мы должны работать непосредственно с (73в), вводя

$$W_{\text{ren}} = -\frac{1}{64\pi^2} \int_0^{\infty} \ln(\varkappa^2 is) \left(\frac{\partial}{i\partial s}\right)^3 \Lambda(s) i ds \quad (m=0, n=4). \tag{82}$$

Здесь с неизбежностью появляется произвольный масштабный множитель $\varkappa$, который следовало ввести с самого начала в определения (62) и (63). Нетрудно убедиться, что во всех предыдущих случаях $\varkappa$ исчезнет из теории после перенормировки, но в данном случае он сохраняется в логарифмическом члене. Таким образом, здесь мы сталкиваемся с неоднозначностью. Легко видеть, что изменение $\varkappa$ приводит к изменению в W_{ren}, пропорциональному A_2. Такое изменение может быть компенсировано определенным контрчленом, но это не поможет нам, так как никогда нельзя избавиться полностью от $\varkappa$. За исключением обращения к эксперименту (который находится далеко за пределами достижимого ввиду того, что здесь мы имеем дело с ничтожно малыми эффектами), у нас, по-видимому, нет никакого способа определить, чему должно быть равно значение $\varkappa$.

Эта неоднозначность связана с неперенормируемостью квантовой гравитации. Когда $n=4$, мы вынуждены ввести две неклассические величины уже в однопетлевом приближении — константу взаимодействия, на которую умножается A_2 и которая должна содержаться в S, и масштабный множитель. Как мы уже отметили, эти две величины не независимы: изменение одной из них может быть компенсировано изменением другой, что представляет собой частный случай более общих соотношений, которые в последние годы получили известность как *уравнения ренормгруппы*. Если бы квантовая гравитация была перенормируема, можно было бы выписать конечное число таких уравнений, причем в каждом было бы конечное число членов. В действительности же требуется бесконечное число членов, отражающих все возрастающее число новых типов бесконечностей (и соответственно неоднозначностей), появляющихся в двухпетлевом и более высоких приближениях (см. статью Вейнберга VIII в настоящем сборнике). С точки зрения общепринятой теории возмущений для фиксации теории при этом потребуется бесконечное число экспериментов.

С другой стороны, результаты предварительных исследований [20, 54, 55] дают серьезные основания думать, что если бы нашлись надежные методы приближенного суммирования полного ряда квантовой теории возмущений, то квантовая гравитация оказалась бы *конечной* и все вторичные величины полностью определялись бы постоянной Планка и классическими постоянными G и c [1]).

---

[1]) Эти исследования имели тот недостаток, что производимые в них суммирования, хотя и включают бесконечное число членов, все же всегда носят частичный характер и приводят к результатам, которые калибровочно-инвариантны в лучшем случае лишь до некоторого порядка. Более того, неожиданным образом возникают неоднозначности в процедурах аналитического продолжения, которые необходимо при этом привлечь. Хокинг [49] высказал предположение, что такого рода подходы могут быть принципиально неверными, обосновывая это тем, что в гравитационном вакууме доминирующую роль (как инстантоны) играют черные дыры планковской массы. Однако существование черных дыр не обязательно означает, что нужно отказаться от техники суммирования, поскольку уже давно известно,

В частности, в той мере, в какой однопетлевое приближение (82) дает грубое приближение к точному эффективному действию, масштабный множитель $\varkappa$ должен быть порядка планковской массы $(\hbar c/G)^{1/2}$. С этой точки зрения недостатком выражения (82) можно считать лишь то, что мы пока не знаем, как вычислить коэффициент пропорциональности между $\varkappa$ и $(\hbar c/G)^{1/2}$.

4.5. КОНФОРМНАЯ ИНВАРИАНТНОСТЬ И АНОМАЛИЯ СЛЕДА

Функционально дифференцируя дифференциальное уравнение (39) и пользуясь представлением (59), можно показать, что

$$T^{\mu\nu}(x, s) = -2i(4\pi)^{n/2}(is)^{n/2+1}\sum_{\alpha}\Big\{T^{\mu\nu}(v(\alpha, x)) + \\ + \frac{1}{2}g^{1/2}(x)\,g^{\mu\nu}(x)\,[\lambda(\alpha)+m^2]\,[v(\alpha, x)]^2\Big\}\,e^{-i[\lambda(\alpha)+m^2]s}, \quad (83)$$

где $T^{\mu\nu}(v, (\alpha, x))$ получается из обычного выражения для тензора натяжений подстановкой $v(\alpha, x)$ вместо скалярного поля ϕ. Используя (57), легко убедиться, что ковариантная дивергенция величины в фигурных скобках в (83) обращается в нуль и, следовательно, $T^{\mu\nu}(x, s)_{;\nu}=0$, что, конечно, согласуется с (80). Столь же непосредственно можно получить (см. [8])

$$T_{\mu}{}^{\mu}(x, s) = g^{1/2}(x)\left[n - 2is\left(m^2 + \frac{\partial}{i\partial s}\right)\right][e^{-im^2s}\Lambda(x, x, s)]. \quad (84)$$

Если выражение (84) подставить в (79), (81) или (82), то мы получим, что при $m=0$ и $\xi={}^1/_4(n-2)/(n-1)$ след перенормированного тензора натяжений в отличие от классического тензора натяжений при четных n *не* обращается в нуль. Это явление известно как *аномалия следа*.

Аномалия следа в данной точке зависит от локальной геометрии в ней. В каждом случае ее можно вычислить с помощью выражения (84), а еще проще следующим образом. Заметим, что второе из уравнений (69) дает

$$g_{\mu\nu}(x)\frac{\delta\zeta(z)}{\delta g_{\mu\nu}(x)} = z\zeta(x, z), \quad (85)$$

что метрику Шварцшильда можно построить из линеаризованной теории гравитации на плоском фоне путем суммирования ряда классической теории возмущений при условии, что допускается введение новых координат и определенное аналитическое продолжение после того, как суммирование выполнено (см. [29]). Идеи Хокинга тем не менее ценны тем, что они фокусируют внимание на том способе аналитического продолжения, который оказался полезным в теории черных дыр. Этот способ, возможно, поможет избавиться от неопределенностей, которые появлялись в упомянутых ранних исследованиях, а акцент на черные дыры может указать путь к калибровочно-инвариантным методам суммирования и более точным методам вычисления эффективного лагранжиана.

откуда после дифференцирования по z получаем

$$g_{\mu\nu}(x)\frac{\delta\zeta'(z)}{\delta g_{\mu\nu}(x)}=\zeta(x,\ z)+z\frac{\partial}{\partial z}\zeta(x,\ z). \qquad (86)$$

Заметим далее, что когда W и W_{ren} определены через дзета-функцию, они отличаются друг от друга на величину, пропорциональную $A_{n/2}$, которая, как мы видели, конформно-инвариантна. Поэтому из (72) и (86) следует

$$\langle \mathbf{T}_{\text{ren}}{}^{\mu}_{\mu}(x)\rangle = 2g_{\mu\nu}(x)\frac{\delta W_{\text{ren}}}{\delta g_{\mu\nu}(x)} = 2g_{\mu\nu}(x)\frac{\delta W}{\delta g_{\mu\nu}(x)} =$$
$$= -ig_{\mu\nu}(x)\frac{\delta\zeta'(0)}{\delta g_{\mu\nu}(x)} = -i\zeta(x,0) = \begin{cases} 0 & (n \text{ нечетное}),\\ (4\pi^{-n/2}g^{1/2}(x)a_{n/2}(x,\ x) & (n \text{ четное}). \end{cases} \qquad (87)$$

В однопетлевом приближении скобки $\langle\ \rangle$ в левой части выражения (87) можно убрать. Это следует (как можно увидеть, построив произвольное пространство Фока) из того факта, что матричные элементы следа между любыми двумя ортогональными состояниями обращаются в нуль. Поэтому след перенормированного *оператора* тензора натяжений будет c-числом (единичным оператором, умноженным на число).

В следующем разделе мы увидим, что в ряде важных случаев аномалии следа весьма просто полностью определяют тензор натяжений. К сожалению, в остальных случаях определение $\langle \mathbf{T}^{\mu\nu}_{\text{ren}}\rangle$ требует сложных вычислений, и оно может быть успешным лишь при условии, что имеется достаточно сведений относительно собственных функций оператора F, входящего в полевое уравнение. Более того, с этими собственными функциями еще требуется провести довольно сложные вычисления. Наивные суммы по модам, которые часто пишут для $\langle \mathbf{T}^{\mu\nu}\rangle$, следует рассматривать с крайней осторожностью, так как они расходятся и должны быть видоизменены вычитаниями, относительно которых не всегда ясно, обеспечивают ли они выполнение условия (22). Эта трудность особенно велика при $m=0$. В этом случае полезно иметь альтернативные суммы, которые уже правильно перенормированы. Примером такой суммы является следующее выражение, полученное подстановкой (83) в (81), представлением $\Lambda(x,\ x,\ s)$ в виде суммы по $v(\alpha,\ x)$ и последующим интегрированием по частям:

$$\langle \mathbf{T}^{\mu\nu}_{\text{ren}}(x)\rangle = \int_0^\infty \sum_\alpha \left\{T^{\mu\nu}(v(\alpha,\ x)) + \frac{1}{4}\lambda(\alpha)\,g^{1/2}(x)\,g^{\mu\nu}(x)\,[v(\alpha,\ x)]^2\right\}\times$$
$$\times e^{-i\lambda(\alpha)s}\,ds + \frac{1}{4\,(4\pi^2)}\,g^{1/2}(x)\,g^{\mu\nu}(x)\,a_2(x,\ x) \qquad (88)$$

($m=0$, $n=4$, $\xi={}^1/_6$, пространство-время конформно-плоское или риччи-плоское). Это выражение, в котором явно выступает анома-

лия следа, было проверено прямым вычислением в простых случаях, и есть основания считать, что оно будет полезно во многих более сложных ситуациях. При этом необходима осторожность: порядок суммирования и интегрирования нельзя менять.

4.6. КОНФОРМНО-ПЛОСКОЕ ПРОСТРАНСТВО-ВРЕМЯ

Браун и Кассиди [9] указали, что при $m=0$ и $\xi={}^1/_4(n-2)/(n-1)$ изменение $\langle T^{\mu\nu}_{\text{ren}}\rangle$, возникающее при конформной вариации метрики, полностью определяется аномалией следа. Это следует из легко проверяемого тождества

$$\left[g_{\mu\nu}(x)\frac{\delta}{\delta g_{\nu\nu}(x)},\quad g_{\alpha\beta}(x')\frac{\delta}{\delta g_{\alpha\beta}(x')}\right]=0, \tag{89}$$

которое при применении к W_{ren} дает

$$\delta\langle T_{\text{ren}\,\mu}{}^{\nu}(x)\rangle=g_{\mu\nu}(x)\frac{\delta}{\delta g_{\nu\nu}(x)}\int\langle T_{\text{ren}\,\alpha}{}^{\alpha}(x')\rangle\,\delta\lambda(x')\,d^nx', \tag{90}$$

где δ обозначает результат вариации метрики на величину $\delta g_{\mu\nu}=g_{\mu\nu}\delta\lambda$.

Функциональная производная в правой части (90) берется непосредственным вычислением. Если метрика уже конформно-плоская, можно проинтегрировать полученное вариационное уравнение полностью так же, как в плоском пространстве. В случае $n=2$, записав $g_{\mu\nu}=e^{\lambda}\eta_{\mu\nu}$, где $\eta_{\mu\nu}$ — метрика Минковского, получаем

$$\langle T_{\text{ren}\,\mu}{}^{\nu}\rangle=\langle T_{\text{ren}\,\mu}{}^{\nu}\rangle_{\flat}+\frac{1}{24\pi}\left(\lambda_{,\nu}{}^{\mu}-\delta^{\nu}_{\mu}\lambda_{,\sigma}{}^{\sigma}-\frac{1}{2}\lambda_{,\mu}\lambda_{,}+\frac{1}{4}\delta_{\mu}{}^{\nu}\lambda_{,\sigma}\lambda_{,}{}^{\sigma}\right) \tag{91}$$

$m=0$, $n=2$, $\xi=0$, $g_{\mu\nu}=e^{\lambda}\eta_{\mu\nu}$), где $\langle T_{\text{ren}\,\mu}{}^{\nu}\rangle_{\flat}$ — перенормированный тензор натяжений в плоском пространстве-времени, а индексы в правой части поднимаются и опускаются с помощью метрики Минковского.

В случае $n=4$, вспоминая, что тензор Римана и тождества Бианки для конформно-плоского пространства-времени можно целиком выразить через тензор Риччи, прямым вычислением устанавливаем, что интегрирование (90) приводит к соотношению

$$\langle T_{\text{ren}\,\mu}{}^{\nu}\rangle=\langle T_{\text{ren}\,\mu}{}^{\nu}\rangle_{\flat}+\frac{g^{1/2}}{2880\pi^2}\Big[-R_{\mu\alpha}R^{\nu\alpha}+RR_{\mu}{}^{\nu}-$$
$$-\frac{1}{3}R_{;\mu}{}^{\nu}+\delta_{\mu}{}^{\nu}\left(\frac{1}{2}R_{\alpha\beta}R^{\alpha\beta}-\frac{1}{3}R^2+\frac{1}{3}R_{;\alpha}^{\;\;\alpha}\right)\Big] \tag{92}$$

($m=0$, $n=4$, $\xi={}^1/_6$, пространство-время конформно-плоское).

При нечетном n мы, естественно, получаем

$$\langle T_{\text{ren}\,\mu}{}^{\nu}\rangle=\langle T_{\text{ren}\,\mu}{}^{\nu}\rangle_{\flat} \tag{93}$$

($m=0$, n нечетно, $\xi={}^1/_4(n-2)/(n-1)$, пространство-время конформно-плоское).

Заметим, что $\langle T^{\nu}_{\mathrm{ren}\ \mu}\rangle_b$ *не обязательно должно быть равно нулю.* Как мы уже видели ранее, это среднее, вообще говоря, зависит от топологии, а если опорное плоское многообразие неполное, то также и от краевых условий. В рассматриваемом случае топология и краевая структура плоского многообразия определяются его конформной связью с искривленным многообразием. Предполагается, что представление (35) для фейнмановского пропагатора при конформных преобразованиях остается неизменным, поэтому проинтегрированные выражения (91) — (93) наследуют определение вакуума от плоского многообразия. Отметим, что эти выражения чисто действительные, откуда следует, что W не имеет мнимой части. Это означает, что рождение частиц отсутствует, а потому $|\mathrm{in, vac}\rangle$ и $|\mathrm{out, vac}\rangle$ идентичны.

Поскольку каждое двумерное многообразие локально конформно-плоское, выражение (91) имеет универсальный характер. Отображая конформно многообразия с искривленными краями на многообразия с плоскими краями, можно выразить тензор натяжений для сложных случаев (например, для двумерного клина Риндлера) через тензор натяжений для простых случаев (например, для полупространства). Эти результаты, как оказалось, согласуются с результатами, полученными иными методами, даже в тех случаях, *когда тензор Римана равен нулю.* Это означает, что конформная аномалия не является следствием кривизны, а представляет собой глубокое внутреннее свойство теории.

Многообразия Робертсона — Уокера составляют важный класс, к которому можно применить равенство (92). Простейшим в этом классе является пространство-время де Ситтера, тензор Римана которого имеет вид $R_{\mu\nu\sigma\tau}=K(g_{\mu\sigma}g_{\nu\tau}-g_{\mu\tau}g_{\mu\sigma})$. Подстановка его в (92) и наложение граничного условия $\langle T^{\mu}_{\mathrm{ren}}\rangle_K \to 0$ немедленно дают

$$\langle T^{\mu\nu}_{\mathrm{ren}}\rangle = -\frac{K^2}{960\pi^2}\, g^{1/2} g^{\mu\nu} \tag{94}$$

($m=0$, $\xi=1/6$, пространство-время де Ситтера), т. е. результат, который был получен с затратой большого труда рядом других методов. Связанное с этим результатом «вакуумное» состояние де ситтер-инвариантно.

5. ПОЛНАЯ КВАНТОВАЯ ТЕОРИЯ

Методы предыдущих разделов применимы к полям со спином в той же мере, в какой они применимы к скалярному полю. Требуется лишь ввести несколько новых технических средств, таких, как битензоры или биспиноры для описания переноса поляризации полей вдоль геодезических, а если имеются калибровочные группы, то ввести также члены, фиксирующие калибровку и «ду́ховы» по-

ля [1]). Но в основном однопетлевая теория для $\langle T^{\mu\nu}_{ren} \rangle$ одинакова как для скалярного поля, так и для электромагнитного поля, поля спина $^1/_2$, поля Прока и гравитационного поля [2]).

Как мы отмечали выше, однопетлевая теория является ВКБ-приближением к полной теории, к которой мы теперь обращаемся. Начиная с этого момента мы сможем делать мало строгих утверждений. Большую часть времени мы будем иметь дело с формальными уравнениями, из которых никто пока не может извлечь надежные конечные ответы. Однако исходным пунктом этих уравнений является функциональный интеграл Фейнмана, в отношении которого широко распространена уверенность, что его справедливость выходит за пределы теории возмущений. Поэтому ожидается, что сами эти уравнения сохранятся при дальнейшем развитии теории.

Для простоты мы рассмотрим чистую квантовую гравитацию, но введем компактную систему обозначений, которая в равной степени применима ко всем полевым теориям. Начнем с замены символа $g_{\mu\nu}(x)$ для квантового метрического тензора символом φ^i. Подразумевается, что индекс i заменяет смешанный набор нумерующих величин μ, ν, x, где x нумерует пространственно-временную точку. Причина включения в этот набор непрерывного «номера» x состоит в том, что большая часть формализма квантовой теории поля носит чисто комбинаторный характер, причем суммирование по немым тензорным индексам часто сопровождается интегрированием по пространству-времени. Для того чтобы избежать написания множества знаков интегрирования, мы объединяем индексы x и тензорные индексы в один индекс и уславливаемся, что по повторяющимся нижним строчным латинским индексам производится комбинированное суммирование-интегрирование. Расширяя множество значений, которое пробегает индекс i, можно включить в символ φ^i также компоненты других полей. Если некоторые из этих полей фермионные, то необходимы дополнительные обозначения, чтобы контролировать знаки и расстановку множителей, но пока мы не будем рассматривать подобное усложнение.

5.1. КАЛИБРОВОЧНАЯ ГРУППА

Единственный другой тип индекса, который будет введен,— это нижний греческий индекс из начала алфавита, называемый *групповым* индексом. Рассмотрим диффеоморфизм, который порождается смещением пространственно-временны́х точек на вариантный вектор $-\delta\xi$. На координатном языке соответствующее преобразование можно записать в виде $\bar{x}^\mu = x^\mu + \delta\xi^\mu$, а изменение $g_{\mu\nu}(x)$ в резуль-

---

[1]) Как они появляются в теории, можно увидеть из полной теории, описанной в последующих разделах.

[2]) В теориях супергравитации имеется также безмассовое поле спина $^3/_2$.

тате такого преобразования имеет вид

$$\delta g_{\mu\nu}(x) = -\pounds_{\delta\xi} g_{\mu\nu}(x) = \int [-g_{\mu\nu,\sigma}(x)\,\delta(x,\ x') - g_{\mu\nu}(x)\,\delta_{,\mu}(x,\ x') - g_{\mu\sigma}\delta_{,\nu}(x,\ x')]\,\delta\xi^{\sigma}(x')\,d^n x'. \quad (95)$$

Преобразование (95) называется *калибровочным преобразованием*, а соответствующая группа диффеоморфизмов называется *калибровочной группой* квантовой гравитации.

Нетрудно показать, что величина в квадратных скобках под интегралом в (95) является битензорной плотностью, преобразующейся как ковариантный тензор в точке x и как ковариантная векторная плотность веса 1 в точке x'. В компактных обозначениях мы заменим эту величину символом $Q^i_{\alpha}[\varphi]$ и перепишем (95) в виде

$$\delta\varphi^i = Q^i{}_{\alpha}[\varphi]\,\delta\xi^{\alpha}. \quad (96)$$

Здесь индексы μ, ν, x заменены индексом i, а индексы σ, x' заменены *групповым* индексом α.

Уравнение (96) выражает действие упомянутого выше диффеоморфизма в пространстве метрических тензоров φ^i. Это действие является *реализацией* диффеоморфизма. Величины Q удовлетворяют важному тождеству, возникающему в силу того факта, что диффеоморфизмы образуют группу

$$Q^i{}_{\alpha,\,j}Q^j{}_{\beta} - Q^i{}_{\beta,\,j}Q^j{}_{\alpha} = Q^i{}_{\gamma}c^{\gamma}{}_{\alpha\beta}. \quad (97)$$

Здесь (и далее везде) запятыми обозначается *функциональное* дифференцирование по φ. Постоянные c являются структурными константами калибровочной группы и удовлетворяют циклическому тождеству [1])

$$c^{\alpha}{}_{\beta\varepsilon}c^{\varepsilon}{}_{\gamma\delta} + c^{\alpha}{}_{\gamma\varepsilon}c^{\varepsilon}{}_{\delta\beta} + c^{\alpha}{}_{\delta\varepsilon}c^{\varepsilon}{}_{\beta\gamma} = 0. \quad (98)$$

Очевидно, что множество значений, которые пробегают групповые индексы, можно расширить так, чтобы включить в рассмотрение другие калибровочные группы, ассоциированные с другими полями (например, с полями Янга — Миллса). Компактные обозначения, в частности в равенствах (97) и (98), остаются прежними. Допуская в отношении структурных констант антикоммутативность и обобщенные симметрии, можно включить сюда даже суперкалибровочные группы.

[1]) В подробных обозначениях структурные константы группы диффеоморфизмов записывают как $c^{\mu}{}_{\nu'\sigma''}$ и определяются соотношением

$$\int d^n x' \int d^n x\, c^{\mu}{}_{\nu'\sigma''}X^{\nu'}Y^{\sigma''} = -[X,\ Y]^{\mu},$$

где X и Y — произвольные постоянные векторные поля, а [,] — скобки Ли; $c^{\mu}{}_{\nu'\sigma''}$ — компоненты трехточечной тензорной плотности, преобразующиеся как контравариантный вектор в точке x и как ковариантная векторная плотность веса 1 в каждой из точек x' и x''. Циклическое тождество, которому они удовлетворяют, есть тождество Якоби для скобок Ли.

Замечательным и часто используемым фактом является то, что поля, которые встречаются на практике, осуществляют *линейную* реализацию соответствующих калибровочных групп. Это означает, что функциональные производные $Q^i_{\alpha,\,j}$ не зависят от φ^i и, если их рассматривать как непрерывные матрицы (по i и j), дают матричное *представление* алгебры Ли, ассоциированной с группой. Конечно, эта простота, вообще говоря, теряется, если φ заменяются их нелинейными функциями. Но замечательно то, что, как правило, имеется «естественный» набор полевых переменных, для которых величины Q являются линейными функционалами [1]).

5.2. ФУНКЦИОНАЛЬНЫЙ ИНТЕГРАЛ ФЕЙНМАНА; МЕРА

Рассмотрим теперь амплитуду перехода вида $\langle \mathrm{out}|\mathrm{in}\rangle$, где векторы $|\mathrm{in}\rangle$ и $|\mathrm{out}\rangle$ относятся к состояниям, в которых поле максимально задано (в квантовомеханическом смысле, например через полный набор коммутирующих наблюдаемых) в областях «in» и «out» соответственно. Эти состояния не обязательно «вакуумные», а области «in» и «out» не обязательно находятся соответственно в бесконечно удаленном прошлом или будущем. Если пространство-время имеет геометрические сингулярности в прошлом и (или) будущем, то $|\mathrm{in}\rangle$ и $|\mathrm{out}\rangle$ могут быть определены не через наблюдаемые, а путем некоторой специальной процедуры аналитического продолжения. Предполагается лишь, что области «in» и «out» лежат соответственно в прошлом или будущем области, представляющей интерес с точки зрения динамики.

Есть много способов показать, что амплитуда $\langle \mathrm{out}|\mathrm{in}\rangle$ может быть выражена в виде формального функционального интеграла:

$$\langle \mathrm{out}\,|\,\mathrm{in}\rangle = N \int e^{iS[\varphi]}\mu[\varphi]\,d\varphi. \tag{99}$$

Здесь N — нормировочная постоянная, $S[\varphi]$ — классический функционал действия, $\mu[\varphi]$ — некоторый функционал меры, а интегрирование распространяется как на все поля φ, которые удовлетворяют граничным условиям, подходящим для данных состояний «in» и «out», так и на все топологии, к которым можно прийти аналитическим продолжением данной фоновой топологии.

Выражение (99) впервые было выведено Фейнманом [37] в обычной квантовой механике, без калибровочных групп, а позднее было применено им к теории поля [38]. Распространение на калибровочные теории — плод труда большого числа исследователей; подробности можно найти в литературе [2]). В сочетании с соотношением

$$\langle \mathrm{out}\,|\,T(A[\varphi])\,|\,\mathrm{in}\rangle = N \int A[\varphi]\,e^{iS[\varphi]}\mu[\varphi]\,d\varphi \tag{100}$$

---

[1]) В теории гравитации имеется семейство «естественных» полей, а именно всех тензорных плотностей вида $g^r g^{\mu\nu}$ или $g^{-r} g^{\mu\nu}$, где $r \neq 1/n$.

[2]) Из современных работ полезны [1, 33, 34].

выражение (99) немедленно приводит к таким вариационным законам, как (1), (2), (29) и (42).

Правильный выбор функционала меры $\mu[\varphi]$ был предметом разногласий в течение многих лет. Нет ни одного хорошего обзора по возникающим при этом проблемам, и размеры данной статьи не позволяют дать такой обзор здесь. Достаточно сказать, что выбор, которому следует тот или иной автор, зависит от следующих факторов: 1) какому определению функционального интеграла отдается предпочтение (например, через разбиение времени или через функции Грина и правила Фейнмана), 2) какой используется формализм — лагранжев или гамильтонов, 3) какая интерпретация принимается для некоторых неопределенных выражений. Многие из этих неопределенностей связаны с проблемой упорядочения сомножителей и с вопросом, какой смысл придать локальным операторам, построенным из исходных полей φ^i. В случае перенормируемых теорий такие неопределенности при адекватной регуляризации исчезают и не появляется никаких разногласий относительно того, как производить вычисления. В случае неперенормируемых теорий, напротив, из-за того, что не существует надежных конечных ответов, противоборствующие интерпретации в настоящее время напоминают средневековые споры о том, сколько чертей уместится на острие иглы.

Мы займем здесь бескомпромиссно формальную позицию, поскольку представляется весьма вероятным, что она останется справедливой в процессе дальнейшего развития, и будем полагаться на известную тенденцию, согласно которой чистый формализм приобретает и сохраняет свою последовательность и логику. Мы начнем с того, что амплитуда (99) должна быть калибровочно-инвариантной. Поскольку функционал действия в экспоненте подынтегрального выражения уже калибровочно-инвариантен, мы обеспечим калибровочную инвариантность интеграла, если потребуем, чтобы «элемент объема» $\mu d\varphi$ был калибровочно-инвариантен. Наиболее простой путь — введение такого метрического тензора γ_{ij} в пространстве полей φ^i, для которого действие калибровочной группы будет изометрией. В явном виде это требование означает

$$0 = \pounds_{Q_\alpha}\gamma_{ij} = \gamma_{ij,\,k}Q^k{}_\alpha + \gamma_{kj}Q^k{}_{\alpha,\,i} + \gamma_{ik}Q^k{}_{\alpha,\,j}. \tag{101}$$

Затем выберем

$$\mu = \text{const} \times [\det(\gamma_{ij})]^{1/2}, \tag{102}$$

предполагая, что детерминант может быть подходящим образом определен. Нетрудно убедиться, что равенство (97) является условием интегрируемости для (101) и что из самого условия (101) формально следует

$$(\mu Q^i{}_\alpha)_{,i} = 0. \tag{103}$$

Это уравнение «потока без расходимости», которое в принципе можно

использовать для отбора подходящего функционала меры независимо от метрики.

В чистой квантовой гравитации с $g_{\mu\nu}$ в качестве основных переменных условие (101) переходит в утверждение о том, что γ_{ij} должна быть двухточечной функцией, преобразующейся при действии группы диффеоморфизмов подобно симметричной контравариантной тензорной плотности веса 1 в каждой из точек. Среди всевозможных таких битензоров имеется единственное (с точностью до постоянного множителя) однопараметрическое семейство битензоров, которые могут быть охарактеризованы как *локальные*. Они имеют вид

$$\gamma^{\mu\nu\sigma'\tau'} = \mathfrak{G}^{\mu\nu\sigma\tau}\delta(x, x'), \tag{104}$$

$$\mathfrak{G}^{\mu\nu\sigma\tau} \overset{\text{def}}{=} \frac{1}{2} g^{1/2}(g^{\mu\sigma}g^{\nu\tau} + g^{\mu\tau}g^{\nu\tau} + \lambda g^{\mu\nu}g^{\sigma\tau}), \quad \lambda \neq -\frac{2}{n}. \tag{105}$$

Из-за дельта-функции, появляющейся в (104), $\gamma^{\mu\nu\sigma'\tau'}$ имеет блочную структуру, что позволяет представить формально его детерминант в виде

$$\det(\gamma^{\mu\nu\sigma'\tau'}) = \prod \mathfrak{G}(x), \tag{106}$$

где $\mathfrak{G}(x)$ — детерминант $^1/_2 n(n+1) \times {^1/_2} n(n+1)$-матрицы $\mathfrak{G}^{\mu\nu\sigma\tau}$. Нетрудно вычислить, что

$$\mathfrak{G} = (-1)^{n-1}\left(1 + \frac{n\lambda}{2}\right) g^{1/4(n-4)(n+1)}, \tag{107}$$

где $g = -\det(g_{\mu\nu})$. В четырехмерном пространстве $\mathfrak{G}$ и, следовательно, $\det(\gamma^{\mu\nu\sigma'\tau'})$, как легко видеть, являются постоянными, независимыми от $g_{\mu\nu}$. Поэтому меру μ можно взять постоянной; без потери общности ее можно приравнять единице. Но это не будет справедливо при иных размерностях или при наличии других полей, если мы будем по-прежнему рассматривать $g_{\mu\nu}$ как основные переменные. Однако в принципе мы можем заменить $g_{\mu\nu}$ переменными $g^{-r}g_{\mu\nu}$ или $g^r g_{\mu\nu}$ ($r \neq 1/n$) и выбрать r так, чтобы μ оставалась постоянной. На практике, как мы увидим ниже, это излишне. Оказывается, что для эффективного приравнивания μ единице достаточно выбрать такие основные поля, которые под действием калибровочной группы преобразуются линейно. Уравнение (99) тогда приобретает простой вид:

$$\langle \text{out} | \text{in} \rangle = N \int e^{iS[\varphi]} d\varphi. \tag{108}$$

5.3. ГЛОБАЛЬНО ВЫПОЛНЯЮЩИЕСЯ КАЛИБРОВОЧНЫЕ УСЛОВИЯ; КАНОНИЧЕСКИЕ КООРДИНАТЫ

Когда есть калибровочная группа, интегрирование в формуле (108) излишне. Поскольку φ пробегает групповую *орбиту* в пространстве полей, экспонента в подынтегральном выражении оста-

ется постоянной [1]). Эту избыточность полезно устранить путем наложения калибровочного условия. Для того чтобы найти калибровочное условие, которое было бы справедливо и в контексте, отличном от теории возмущений, необходимо проявлять осторожность.

Обозначим пространство всех полей φ^i, по которым берется интеграл (108), через Φ, а калибровочную группу — через G. В чистой теории гравитации Φ есть Riem (M) — пространство всех римановых метрик на пространственно-временно́м многообразии M, а G есть Diff (M) — группа диффеоморфизмов M. *Физическое* пространство полей есть фактор-пространство Φ, называемое также *пространством орбит*. В теории гравитации пространство орбит Riem (M)/Diff(M) есть пространство *геометрий* на M.

Рассмотрим типичную (т. е. *общего вида*) орбиту. По модулю возможного дискретного центра она является *копией* группового многообразия M, поскольку дает реализацию G и обладает той же размерностью. Не обязательно, чтобы все орбиты имели именно эту размерность. Часто существует класс вырожденных орбит, имеющих меньшую размерность. Такие орбиты являются *граничными точками* пространства Φ/G и порождаются теми точками в Φ, которые остаются инвариантными под действием нетривиальных (непрерывных) подгрупп группы G. Чем больше размерность такой подгруппы инвариантности, тем меньше размерность орбиты. Фишер [40] показал, что орбиты данной размерности могут быть собраны в граничные подмногообразия, так что все пространство орбит становится слоистым (stratified) многообразием. В теории гравитации вырожденными орбитами являются симметричные геометрии, т. е. те, которые обладают векторами Киллинга [2]). Для наглядности можно представлять себе Φ как R^3, а G — как группу вращений вокруг фиксированной оси. Орбитами тогда будут окружности, перепендикулярные оси и с центром на ней, а пространством орбит будет полуплоскость, граничные точки которой соответствуют точкам на оси, остающимся инвариантными под действием рассматриваемой группы вращений.

Глобально выполняющиеся калибровочные условия представляют собой набор связей, выделяющий в Φ подпространство коразмерности, равной размерности G, которое пересекает каждую орбиту точно в одной точке. В случае поля Янга — Миллса такого подпространства *не существует*, если калибровочная группа соответствует скрученному (twisted) расслоенному пространству. Относительно группы диффеоморфизмов известно значительно меньше. Хотя ее

---

[1]) Калибровочная инвариантность действия $S[\varphi]$ может быть выражена в виде $S_{,i}Q'_{\alpha}\equiv 0$; в чисто гравитационной теории это — свернутое тождество Бианки.

[2]) Строго говоря, это верно только в том случае, если пространство-время компактно или имеет компактные пространственные сечения. Пространство Минковского, например, не является вырожденной орбитой, поскольку от калибровочных параметров $\delta\xi^\sigma$ в (95) требуется, чтобы они обращались в нуль на бесконечности. Поэтому изометрии Пуанкаре не содержатся в калибровочной группе.

структура зависит от природы пространственно-временнóго многообразия M, ее нельзя рассматривать как расслоенное пространство. Можно привести правдоподобные соображения в пользу предположения, что в случае диффеоморфизмов для широкого круга топологий M *можно* сформулировать глобально выполняющиеся калибровочные условия. Здесь мы просто *допустим*, что подпространство с нужными свойствами существует. Можно считать, что это подпространство *представляет* пространство орбит Φ/G. Каждая орбита представляется точкой, в которой она пересекает данное подпространство. Для того чтобы выразить эту мысль в уравнениях, можно представить себе, что переменные φ^i заменены другими переменными $I^A[\varphi]$, $P^\alpha[\varphi]$, где переменная I^A нумерует орбиты и является калибровочно-инвариантной, а переменная P^α нумерует соответственные точки *на* каждой орбите. Точку на каждой орбите, выделенную данным калибровочным условием, можно принять за начало «координат» P^α на этой орбите. Тогда калибровочное условие имеет простой вид: $P^\alpha=0$.

В действительности оказывается более удобным работать с континуумом калибровочных условий

$$P^\alpha[\varphi]=\zeta^\alpha, \tag{109}$$

где ζ^α— постоянные (т. е. не зависят от φ^i), имеющие значения от $-\infty$ до ∞. Чтобы включать всю эту область значений, P^α должны представлять собой весьма специальную систему координат на орбитах. Эти координаты (в принципе) могут быть введены следующим образом. Вспоминая, что каждая орбита (общего вида) является копией G, положим, что начало $P^\alpha=0$ на каждой орбите отождествлено с единичным элементом группы G. Затем выберем в качестве переменных P^α *канонические* координаты группы, т. е. нормальные координаты, порожденные однопараметрическими абелевыми подгруппами G. Имея в виду, что действие калибровочной группы на каждой орбите (общего вида) имитирует ее действие в самой себе, мы приходим к функциональному дифференциальному уравнению

$$P^\alpha{}_{,i}[\varphi]\,Q^i{}_\beta[\varphi]=Q^\alpha{}_\beta[P[\varphi]], \tag{110}$$

где $Q^\alpha{}_\beta$ находятся в том же отношении к групповому многообразию, в каком $Q^i{}_\beta$ находятся к Φ. В частности, они удовлетворяют соотношению, аналогичному соотношению (97):

$$Q^\alpha{}_{\beta,\delta}Q^\delta{}_\gamma-Q^\alpha{}_{\gamma,\delta}Q^\delta{}_\beta=Q^\alpha{}_\delta c^\delta{}_{\alpha\beta}, \tag{111}$$

где греческие индексы после запятой обозначают функциональные производные по P^α. Легко убедиться, что соотношение (111) является условием интегрируемости для уравнения (110).

Уравнения (110) и (111) справедливы при любой координатизации группы. Канонические же координаты выделены тем видом, который принимают при их использовании величины $Q_\beta{}^\alpha$. Можно

показать [21], что

$$Q^{-1}[P] = \frac{e^{c\cdot P} - I}{e^{c\cdot P}} \overset{\text{def}}{=} I + \frac{1}{2!} c\cdot P + \frac{1}{3!}(c\cdot P)^2 + \ldots, \qquad (112)$$

где I — единичная матрица (дельта-функция) и

$$Q(P) \overset{\text{def}}{=} [Q^{\alpha}{}_{\beta}[P]], \quad c\cdot P \overset{\text{def}}{=} (c^{\alpha}{}_{\gamma\beta} P^{\gamma}). \qquad (113)$$

Ряд (112) сходится для *всех* значений P^{α}. Для некоторых значений (непрерывная) матрица Q^{-1} может иметь нулевые корни, что вызовет неограниченный рост правой части равенства (110). При таких значениях система канонических координат становится сингулярной. Однако сингулярности не являются пороком этой системы. Канонические P^{α} аналогичны угловым координатам. По мере пробегания ими допустимой области (от $-\infty$ до ∞) каждая орбита может быть пройдена много (бесконечное число) раз. Но каждый набор значений P^{α} на данной орбите по-прежнему задает единственную точку на ней.

Для полного определения функционалов $P^{\alpha}[\varphi]$ недостаточно одних только уравнений (110). Нужны дополнительные условия, чтобы «выровнять» соответствующие точки на соседних орбитах. Один из возможных способов такого выравнивания состоит в следующем. Введем в пространстве полей Φ несингулярную метрику γ_{ij}, удовлетворяющую требованию (101). Потребуем, чтобы матрица $\gamma_{ij} Q^{i}_{\alpha} Q^{j}_{\beta}$ была несингулярной на всех орбитах общего вида, чтобы никакой вектор не мог быть одновременно и касательным, и ортогональным к какой-либо из орбит. Тогда в пространстве орбит Φ/G может быть индуцирована ассоциированная метрика путем задания расстояния между соседними орбитами в Φ/G как нормального расстояния между ними в Φ. В чисто гравитационной теории если γ_{ij} берется в виде (104), (105) и выбирается $\lambda = -1$, то $\gamma_{ij} Q^{i}{}_{\alpha} Q^{j}{}_{\beta}$ представляет собой некоторое обобщение оператора Лапласа — Бельтрами для векторных полей. Для асимптотически плоского пространства-времени этот оператор несингулярен, так что требуемое условие может встретиться в практически важных случаях. Другое свойство этой частной метрики γ_{ij} состоит в том, что с ее помощью любая пара точек в Φ может быть соединена единственной геодезической. Методы доказательства этого свойства можно найти в работе [22], и в дальнейшем мы будем предполагать наличие такого свойства.

Выберем некоторую орбиту общего вида и назовем ее *базисной орбитой*. Единицу группы на этой орбите назовем *базисной точкой*. Пусть V есть подпространство пространства Φ, порожденное множеством всех геодезических, выходящих из базисной точки в направлениях, ортогональных к базисной орбите. Можно показать [22], что геодезическая, пересекающая ортогонально одну орбиту,

ортогональна ко всем орбитам, которые она пересекает на своем протяжении, и, более того, она описывает геодезическую кривую в пространстве орбит Φ/G. Используя тот факт, что каждая пара точек в Φ может быть соединена единственной геодезической и геодезическая не может быть одновременно и касательной, и ортогональной к какой-либо орбите, можно показать, что V непременно пересекает все орбиты. Чтобы избежать пересечения данной орбиты подпространством V более одного раза, мы можем оборвать каждую из порождающих геодезических, как только она коснется граничной точки пространства Φ/G. Тогда V есть топологическая (но не обязательно метрическая) копия Φ/G [1]).

Если построить другое подпространство V', подобное V, но выходящее из другой точки на базисной орбите, то оно тоже будет пересекать все орбиты. Поскольку групповые операции представляют собой изометрии γ_{ij}, все P^α будут постоянны на V'. Другими словами, если единичные точки «выровнены», то все остальные точки тоже автоматически будут выровнены.

Использование ортогональности для построения представляющего подпространства Φ/G — не новая идея. Однако она обычно применяется в довольно несовершенной форме путем выбора функционалов P^α в виде

$$P^\alpha[\varphi] = Q^i{}_\alpha[g]\,\gamma_{ij}[g]\,\phi^j, \qquad \phi^i \overset{\text{def}}{=} \varphi^i - g^i, \tag{114}$$

где g^i — фоновое поле. Например, если γ_{ij} имеет вид (104), (105), то при выборе P^α в виде (114) условие $P^\alpha=0$ имеет форму

$$g^{1/2}(2\phi_\mu{}^\nu + \lambda\delta_\mu{}^\nu\phi_\sigma{}^\sigma)_{;\nu} = 0. \tag{115}$$

Если еще положить $\lambda=-1$, то получим калибровочное условие, весьма популярное в теории гравитации. (Здесь поднятие и опускание индексов происходит с помощью фоновой метрики $g_{\mu\nu}$, и через нее же определена ковариантная производная.)

Условие $P^\alpha=0$ с P^α в виде (114) или в более общем случае условие (109) с P^α в виде $P_{\alpha i}[g]\,\phi^i$ при некоторых $P_{\alpha i}$ называется *линейным* калибровочным условием. Линейные калибровочные условия весьма удобны в теории возмущений, где квантованное поле никогда не отличается слишком сильно от фонового поля. Ковариантные (по отношению к фоновому полю) калибровочные условия, подобные (115), обычно являются наилучшими, но для некоторых целей более полезны нековариантные калибровки (например, кулоновская

---

[1]) Чтобы получить представление о том, какие могут возникнуть метрические ситуации, представим себе Φ как R^3 и G как группу винтовых движений вокруг некоторой оси с фиксированным шагом. Тогда орбитами будут винтовые линии и все они — орбиты общего вида. Если R^3 несет декартову метрику, то пространство орбит топологически (но не метрически) плоское. Отметим, что в этом примере нет поверхностей, которые пересекали бы все орбиты ортогонально, хотя любая плоскость, не содержащая оси, перпендикулярна некоторой орбите в точке пересечения и является поверхностью, подобной V, основанной на этой орбите.

калибровка в теории Янга — Миллса). Но в непертурбативных подходах линейными калибровочными условиями нужно пользоваться с большой осторожностью (см. [43]). Это можно видеть уже из условия типа (115), которое задает подпространство, приближенно ортогональное только тем орбитам, которые лежат достаточно близко к фоновой орбите, проходящей через $\phi_{\mu\nu}=0$. При глобальном применении линейных калибровочных условий могут встретиться по крайней мере следующие пять трудностей. 1. Подпространство, определенное линейным условием, не имеет границы; следовательно, оно не может правильно представлять Φ/G, если есть вырожденные орбиты. 2. Подпространство, определенное линейным условием, может пересекать некоторые орбиты более одного раза. 3. Могут быть орбиты, которые оно вовсе не пересекает. 4. Даже если оно пересекает все орбиты при определенных значениях ζ^α в равенстве (109), оно может не пересекать все орбиты при других значениях ζ^α. 5. Когда G «скручено», вообще нет глобально выполняющихся калибровочных условий, ни линейных, ни каких-либо иных.

В некоторых случаях оказывается возможным сохранить преимущества линейных калибровочных условий. Это достигается в некоторых глобальных подходах к теории Янга — Миллса. Однако группа диффеоморфизмов теории гравитации намного более сложна, чем группа Янга — Миллса, и в настоящее время почти не изучены ни трудности, к которым она приводит глобально, ни предоставляемые ею возможности технических усовершенствований. Для того чтобы сохранить открытыми все пути, мы будем развивать сначала формальную теорию амплитуды $\langle\mathrm{out}|\mathrm{in}\rangle$, используя надежные калибровочные условия, основанные на канонических координатах, а затем сделаем ряд замечаний относительно того, как могли бы обстоять дела, если бы использовались иные калибровочные условия.

5.4. ОТФАКТОРИЗАЦИЯ КАЛИБРОВОЧНОЙ ГРУППЫ

Способ устранения избыточности из интеграла (108) с использованием калибровочного условия впервые был указан Фаддеевым и Поповым [35.] Здесь мы будем придерживаться их метода с некоторыми усовершенствованиями. Пусть ξ — произвольный элемент калибровочной группы (рассматриваемой как абстрактная группа) и пусть ξ^α — канонические координаты ξ. Пусть φ^i — произвольное поле в Φ. Обозначим через ${}^{\xi}\varphi^i$ поле, в которое переводится φ^i под действием ξ. Введем

$$\Delta[\zeta,\ \varphi] \stackrel{\mathrm{def}}{=} \int_G [P[{}^{\xi}\varphi]-\zeta] \det Q^{-1}[\xi]\, d\xi, \quad d\xi \stackrel{\mathrm{def}}{=} \prod_\alpha d\xi^\alpha, \qquad (116)$$

где $\delta[\ldots]$ — дельта-*функционал*, $P^\alpha[\varphi]$ — функционалы, построенные по канонической процедуре, описанной в предыдущем раз-

деле, Q^{-1} — матрица, определенная рядом (112), а интегрирование распространяется на всю калибровочную группу. Из-за присутствия дельта-функционала подынтегральное выражение «срабатывает» только в одной точке группы G, а именно в той точке, для которой ${}^{\xi}\varphi^i$ равно единственному полю φ^i_ζ, которое лежит на орбите, содержащей φ^i, и выделено калибровочным условием (109):

$$P^\alpha[\varphi_\zeta] = \zeta^\alpha. \qquad (117)$$

Определитель $\det Q^{-1}[\xi]$, появляющийся в (116), есть известная правоинвариантная мера на G; он удовлетворяет равенству

$$\det Q^{-1}[\xi\xi']\, d(\xi\xi') = \det Q^{-1}[\xi]\, d\xi \text{ для всех } \xi' \in G, \qquad (118)$$

где $\xi\xi'$— групповое произведение ξ и ξ'. Его присутствие обеспечивает калибровочную инвариантность функционала Δ:

$$\Delta[\zeta, {}^{\xi'}\varphi] = \Delta[\zeta, \varphi] \text{ для всех } \xi' \in G. \qquad (119)$$

Вследствие этой калибровочной инвариантности легко вычислить Δ. Нужно лишь сдвинуть φ^i в φ^i_ζ, чтобы подынтегральное выражение «срабатывало» в единичном элементе $\xi^\alpha = 0$. Теперь все величины могут быть разложены в степенные ряды по ξ^α. Например, аргумент дельта-функционала принимает вид

$$P^\alpha[{}^{\xi}\varphi_\zeta] - \zeta^\alpha = P^\alpha[\varphi_\zeta] - \zeta^\alpha + P^\alpha{}_{,i}[\varphi_\zeta]\, Q^i{}_\beta[\varphi_\zeta]\, \xi^\beta = \\ = Q^\alpha{}_\beta[P[\varphi_\zeta]]\, \xi^\beta + \ldots, \qquad (120)$$

откуда

$$\Delta[\zeta, \varphi] = \int\limits_G \delta[Q[P[\varphi_\zeta]]\xi + \ldots]\left(1 + \frac{1}{2}\operatorname{tr} c\cdot\xi + \ldots\right) d\xi = \\ = (\det F[\varphi_\zeta])^{-1}, \qquad (121)$$

где F — матрица с элементами

$$F^\alpha{}_\beta[\varphi] \overset{\text{def}}{=} Q^\alpha{}_\beta[P[\varphi]]. \qquad (122)$$

В более общем случае, когда P_α не выбраны так, чтобы удовлетворять дифференциальному уравнению (110), F определяется выражением

$$F^\alpha{}_\beta[\varphi] \overset{\text{def}}{=} P^\alpha{}_{,i}[\varphi]\, Q^i{}_\beta[\varphi]. \qquad (123)$$

Теперь подставим в подынтегральное выражение (108) единицу в виде $(\Delta[\zeta, \varphi])^{-1} \int\limits_G \delta[P[{}^{\xi}\varphi] - \zeta] \det Q^{-1}[\xi]\, d\xi$ и поменяем порядок интегрирования; это дает

$$\langle \text{out} | \text{in} \rangle = N \int\limits_G \det Q^{-1}[\xi] d\xi \int d\varphi\, e^{iS[\varphi]} (\Delta[\zeta, \varphi])^{-1} \delta[P[{}^{\xi}\varphi] - \zeta]. \qquad (124)$$

Мы видели выше, что элемент объема $d\varphi$ калибровочно-инвариантен. (Напомним, что теперь $\mu = \text{const}$.) $S[\varphi]$ и $\Delta[\zeta, \varphi]$ также калибровоч-

но-инвариантны. Поэтому верхний индекс ξ может быть приписан любому φ в подынтегральном выражении (124), если только данное φ не имеет уже его. Но тогда любое ${}^{\xi}\varphi$ есть немое поле, и все ξ можно опустить. Используя (121), мы немедленно получаем

$$\langle\text{out}|\text{in}\rangle = N' \int e^{iS[\varphi]} \det F[\varphi]\, \delta[P[\varphi]-\zeta]\, d\varphi, \qquad (125)$$

где под интегралом $F[\varphi_\zeta]$ заменено на $F[\varphi]$ из-за присутствия дельта-функционала и где

$$N' \overset{\text{def}}{=} N \int_G \det Q^{-1}[\xi]\, d\xi. \qquad (126)$$

Таким образом, калибровочная группа отфакторизована и ее «объем» учтен в новой нормировочной постоянной N'. Интегрирование в (125) ограничивается подпространством $P^\alpha[\varphi]=\zeta^\alpha$.

Если P^α не выбраны так, чтобы удовлетворять уравнению (110) (например, если мы следуем линейному калибровочному уравнению), может понадобиться видоизменить выражение (125). Например, если подпространство $P^\alpha[\varphi]=\zeta^\alpha$ пересекает определенные орбиты более одного раза, то в подынтегральное выражение необходимо ввести множитель $1/n[\zeta, \varphi]$, где $n[\zeta, \varphi]$ — число пересечений орбиты, содержащей φ^i. Намного более сложная проблема возникает в случае, когда некоторые из орбит вообще не пересечены. Если φ оказывается на одной из таких орбит, то $\Delta[\zeta, \varphi]$ обращается в нуль и выражение (124) теряет смысл. На границе между теми орбитами, которые пересекаются, и теми, которые не пересекаются, $\Delta[\zeta, \varphi]$ обычно имеет поведение типа точки ветвления, и матрица $F[\varphi_\zeta]$ в этой области неограниченно растет. Возможно, что какая-то процедура аналитического продолжения дает возможность обойти эту точку ветвления, особенно если дельта-функционал в (125) представить в виде интеграла Фурье. Но это лишь программа для будущих исследований.

Метод ограничения полей φ^i конкретным подпространством можно применить также для доопределения операторов. Рассмотрим выражение (100). Оно строго справедливо лишь при условии, что функционал $A[\varphi]$ калибровочно-инвариантен. Если это не так, интеграл не определен, подобно интегралу $\int_{-\infty}^{\infty} x\,dx$. Иначе говоря, матричные элементы можно определить только для калибровочно-инвариантных операторов. Однако из заданного калибровочно-неинвариантного оператора $A[\varphi]$ можно построить калибровочно-инвариантный оператор, определяя его следующим образом:

$$T(A[\varphi_\zeta]) \overset{\text{def}}{=} T\left((\Delta[\zeta, \varphi])^{-1} \int_G A[{}^{\xi}\varphi]\, \delta[P[{}^{\xi}\varphi]-\zeta] \det Q^{-1}[\xi]\, d\xi\right). \qquad (127)$$

Здесь использован символ хронологического упорядочения, так что можно пренебречь некоммутативностью оператора $A[{}^{\xi}\varphi]$ как с $(\Delta[\zeta, \varphi])^{-1}$, так и с дельта-функционалом. Отметим, что из-за линейности действия калибровочной группы на поля φ, не возникает никакой неопределенности с символом ${}^{\xi}\varphi$. Вместе с тем следует иметь в виду, что в теории гравитации диффеоморфизмы могут переставлять поле весьма сложным образом. Соответственно хронологическое упорядочение, которое расставляет полевые операторы по значению одной только координаты x^0, постоянно сложным образом перестраивает «физические» поля по мере того, как ξ в интеграле (127) пробегает по группе.

Подставляя (127) в матричный элемент (100) и следуя тем же рассуждениям, что и при переходе от (124) к (125), находим, что равенство

$$\langle \mathrm{out} | T(A[\varphi_\zeta]) | \mathrm{in} \rangle = N' \int A[\varphi] e^{iS[\varphi]} \det F[\varphi] \, \delta[P[\varphi] - \zeta] \, d\varphi \quad (128)$$

справедливо для любого функционала $A[\varphi]$.

5.5. ГАУССОВЫ СРЕДНИЕ

Теорию возмущений можно развить на основе (125) и (128), но обычно более удобно работать с формализмом, из которого дельта-функционалы уже исключены. Отметим, что, хотя в выражение (125) входят параметры ζ^α, в действительности амплитуда $\langle \mathrm{out} | \mathrm{in} \rangle$ не зависит от них. Поэтому ничто не изменится, если проинтегрировать по этим параметрам с некоторым весовым множителем. Гауссов весовой множитель оказывается для этого самым удобным.

Пусть $\gamma_{\alpha\beta}$ — произвольная несингулярная непрерывная матрица; обычно $\gamma_{\alpha\beta}$ выбирается локальной (т. е. пропорциональной дельта-функционалу) и ковариантно зависящей от фонового поля g^i (g^i может быть, например, базисной точкой порожденного геодезическими подпространства V, которая служит нулевой точкой для функционалов P^α, когда те выбираются каноническими). Не выписывая явно индексы, можно формально написать

$$\int e^{\frac{1}{2} i \zeta \gamma \zeta} d\zeta = C (\det \gamma)^{-1/2}, \qquad d\zeta \overset{\mathrm{def}}{=} \prod_\alpha \delta\zeta^\alpha, \quad (129)$$

где C — (расходящаяся) постоянная. Если в подынтегральное выражение (125) подставлен гауссов весовой множитель $e^{\frac{1}{2} i \zeta \gamma \zeta}$ и проведено интегрирование по ζ, то в силу (129) имеем

$$\langle \mathrm{out} | \mathrm{in} \rangle = N'' (\det \gamma)^{1/2} \int e^{i \left(S[\varphi] + \frac{1}{2} P[\varphi] \gamma P[\varphi] \right)} \det F[\varphi] \, d\varphi, \quad (130)$$

$$N'' \overset{\mathrm{def}}{=} N'/C. \quad (131)$$

Этот результат был впервые получен автором через линейные P [23]. Отметим, что при использовании канонических P каждая

орбита при интегрировании может быть пройдена много (бесконечное число) раз.

Выражения (127) и (128) также можно заменить их гауссовыми средними. Вводя

$$T(A[\varphi]) \overset{\text{def}}{=} C^{-1}(\det\gamma)^{1/2}\int T(A[\varphi_\zeta])\, e^{1/2 i\zeta\gamma\zeta}\, d\zeta, \tag{132}$$

можно написать

$$\langle \text{out}\,|\,T(A[\varphi])\,|\,\text{in}\rangle = N''(\det\gamma)^{1/2}\int A[\varphi]\, e^{i\left(S[\varphi]+\frac{1}{2}P[\varphi]\gamma P[\varphi]\right)} \times \det F[\varphi]\, d\varphi. \tag{133}$$

Уравнение (133) и его обобщения часто используются в следующем разделе. Определения (127), (132) дают точное представление о том, какого рода усредненный квантовый оператор ассоциируется в этом формализме с каждым классическим функционалом $A[\varphi]$. Заметим, что

$$T(P^\alpha[\varphi_\zeta]) = \zeta^\alpha, \quad \langle \text{out}\,|\,T(P^\alpha[\varphi])\,|\,\text{in}\rangle = 0. \tag{134}$$

После таких вольных манипуляций с формальными выражениями нам следует теперь убедиться, что в наши результаты не привнесено никаких несоответствий, для чего нужно непосредственно проверить, например, что правая часть (130) в самом деле не зависит от того, как выбраны функционалы $P^\alpha[\varphi]$ и матрицы $\gamma_{\alpha\beta}$. Очевидно, правая часть (130) *будет* затронута, если мы наивно перейдем к P^α, для которых подпространства $P^\alpha = \zeta^\alpha$ не пересекают всех орбит или же пересекают их неоднократно. Однако при инфинитезимальных изменениях P^α и $\gamma_{\alpha\beta}$ должна иметь место инвариантность.

Способ проверки состоит в следующем. Заменим каждое φ^i в интеграле на $\bar{\varphi}^i$, где [1])

$$\bar{\varphi}^i \overset{\text{def}}{=} \varphi^i + Q^i{}_\alpha[\varphi]\,\delta\xi^\alpha[\varphi], \tag{135}$$

$$\delta\xi^\alpha[\varphi] \overset{\text{def}}{=} F^{-1\alpha}{}_\beta[\varphi]\left(\delta P^\beta[\varphi] + \frac{1}{2}\gamma^{-1\beta\gamma}\delta\gamma_{\gamma\delta}P^\delta[\varphi]\right). \tag{136}$$

Поскольку φ^i — переменные интегрирования, их замена остается без последствий. Однако нетрудно показать (подробности см. в работе [23] или [25]), что 1) экспонента получает точно такое изменение, какое она получила бы, если бы P^α и $\gamma_{\alpha\beta}$ были изменены на инфинитезимальные величины δP^α и $\delta\gamma_{\alpha\beta}$ соответственно; 2) то же самое верно для произведения $(\det\gamma)^{1/2}\det F[\varphi]d\varphi$, если считается допустимым положить

$$Q^i{}_{\alpha,\,i} = 0, \quad c^\beta{}_{\alpha\beta} = 0. \tag{137}$$

С такой оговоркой инвариантность интеграла (130) доказана.

---

[1]) При многих способах выбора P матрица F является эффективно дифференциальным оператором и, следовательно, F^{-1} (136) — функцией Грина. Эта функция Грина должна удовлетворять краевым условиям, согласованным с сотояниями «in» и «out».

Уравнения (137) не были нужны для вывода формулы (130). Почему же они понадобились теперь? Ответ заключается в том, что они навязаны нам процедурой отфакторизации калибровочной группы. Перестановка нами порядка интегрирований при получении выражения (124) и использование равенства (126) равнозначны тому, что мы формально обращаемся с калибровочной группой так, как если бы она была *компактной*. При этом для согласованности ассоциированная алгебра Ли также должна рассматриваться как компактная. Генераторы действительных представлений компактных алгебр Ли все должны иметь нулевой след. Отсюда равенства (137).

В теориях Янга — Миллса условия (137) выполняются автоматически, поскольку порождающая группа всегда компактна. В теории гравитации ситуация более сложная. Как $Q^i{}_{\alpha,\,i}$, так и $c^\beta{}_{\alpha\beta}$, если пытаться их вычислить, представляют собой бессмысленные выражения, включающие производные дельта-функционалов с совпадающими аргументами. Однако оба они являются не зависящими от метрики векторными плотностями веса 1. Любая разумная регуляризация *должна* приписать им значение нуль, так как иначе пространство-время будет иметь выделенное направление еще до того, как оно будет наделено метрикой. Можно заметить, что условия (137) и (103) согласуются с условием $\mu=1$. Однако вывод о том, что $Q^i_{\alpha,\,i}=0$, возникает из одного только факта, что под действием группы диффеоморфизмов поля φ^i преобразуются линейно, и никак не связан с конкретным выбором метрики γ_{ij} в виде (104), (105). Это означает, что вклады в функциональный интеграл, возникающие от степеней g, которые сохраняются в определителе (106), (107) при размерностях, отличных от 4, или при выборе полевых переменных, отличных от $g_{\mu\nu}$, должны быть подавлены в любой жизнеспособной схеме регуляризации [1]).

5.6. «ДУ́ХИ»; ПРЕОБРАЗОВАНИЯ БРС; ПРОИЗВОДЯЩИЙ ФУНКЦИОНАЛ

Формулы (130) и (133) ведут к правилам теории возмущений стандартного типа. Присутствие P^α в экспоненте нарушает калибровочную симметрию и устраняет избыточность, имеющуюся в интеграле (108). Но одна важная симметрия остается. Легче всего увидеть ее, вводя два новых *антикоммутирующих* поля χ_α и ψ^α и используя формальные правила интегрирования по таким полям, введенные Березиным [6]. Эти правила аналогичны во многих отношениях известным правилам для обычных определенных интегралов от $-\infty$ до ∞ с подынтегральными выражениями, асимптотически обращающимися в нуль. Например, интегралы от полных производных обращаются в нуль и положение нулевой точки можно

[1]) Сохраняющиеся степени g дают в экспоненту функционального интеграла вклад вида $\mathrm{const}\times\delta(0)\int \ln g\,d^nx$.

сдвигать. С другой стороны, при использовании правил Березина преобразования переменных и взятие гауссовых интегралов ведут к детерминантам, в точности обратным тем, которые дает стандартная теория. В частности, получаем

$$\int \exp(i\chi_\alpha F^\alpha{}_\beta \psi^\beta)\, d\chi\, d\psi = C' \det F[\varphi], \tag{138}$$

где C' — некоторая (расходящаяся) постоянная. Этой формулой можно воспользоваться для того, чтобы представить интеграл (130) в виде

$$\langle \mathrm{out} | \mathrm{in} \rangle = \bar{N} (\det \gamma)^{1/2} \int \exp\left[i \left(S[\varphi] + \frac{1}{2} P[\varphi] \gamma P[\varphi] + \right.\right.$$
$$\left.\left. + \chi F[\varphi] \psi \right) \right] d\varphi\, d\chi\, d\psi, \tag{139}$$

$$\bar{N} \overset{\mathrm{def}}{=} N''/C'. \tag{140}$$

Бекки, Руэ и Стора (БРС) [4] обнаружили, что как экспонента, так и элемент объема $d\varphi d\chi d\psi$ в интеграле (139) инвариантны относительно совокупности преобразований, инфинитезимальные формы которых имеют вид [1])

$$\begin{aligned} \delta\varphi^i &= Q^i{}_\alpha [\varphi] \psi^\alpha \delta\lambda, \\ \delta\chi_\alpha &= \gamma_{\alpha\beta} P^\beta [\varphi]\, \delta\lambda, \\ \delta\psi^\alpha &= -\frac{1}{2} c^\alpha{}_{\beta\gamma} \psi^\beta \psi^\gamma \delta\lambda, \end{aligned} \tag{141}$$

где $\delta\lambda$ — инфинитезимальная антикоммутирующая постоянная. Используя антикоммутативность переменных χ, ψ, $\delta\lambda$, с учетом тождества (97) и определения (123) нетрудно убедиться в инвариантности экспоненты. Вычисляя якобиан Березина для преобразований БРС, находим, что элемент объема $d\varphi d\chi d\psi$ также инвариантен при условии, что выполняются уравнения (137). Прямым вычислением можно убедиться, что преобразования БРС образуют абелеву группу, если ограничиться лишь переменными φ и ψ. Включение переменных χ нарушает групповое условие, если только не выполняется равенство $F^\alpha{}_\beta[\varphi]\psi^\beta = 0$.

Поля χ_α и ψ^α называют *«ду́ховыми» полями*. Они не приводят к появлению физических квантов, но играют важную роль в теории, особенно при введении так называемого производящего функционала. Первоначальная теория производящего функционала принадлежит Зинн-Жюстену; мы будем следовать изложению этой теории, данному в работе [60] и примененному к случаю, когда P^α нелинейны.

---

[1]) Преобразования БРС для квантовой гравитации в частных калибровках были указаны в работах [16, 26]. В работе [65] преобразования БРС были применены к квантованию теорий гравитаций с высшими производными.

Начнем с замены экспоненты в формуле (139) на

$$\tilde{S}[\varphi,\ \chi,\ \psi,\ K,\ L,\ M]+J_i\varphi^i+\bar{J}^\alpha\chi_\alpha+\hat{J}_\alpha\psi^\alpha,$$

где

$$\tilde{S}[\varphi,\ \chi,\ \psi,\ K,\ L,\ M]\overset{\text{def}}{=}S[\varphi]+\frac{1}{2}P^\alpha[\varphi]\gamma_{\alpha\beta}P^\beta[\varphi]+\chi_\alpha F^\alpha{}_\beta[\varphi]\psi^\beta+$$
$$+K_iQ^i{}_\alpha[\varphi]\psi^\beta+\frac{1}{2}L_\alpha c^\alpha{}_{\beta\gamma}\psi^\beta\psi^\gamma+MP^\alpha[\varphi]\gamma_{\alpha\beta}F^\beta{}_\gamma[\varphi]\psi^\gamma, \quad (142)$$

и с обобщения выражения (133):

$$\langle\text{out}\,|T\,(A[\varphi,\ \chi,\ \psi])\,|\,\text{in}\rangle\overset{\text{def}}{=}$$
$$\overset{\text{def}}{=}\bar{N}\,(\det\gamma)^{1/2}\int A[\varphi,\ \chi,\ \psi]\,e^{i\,(\tilde{S}+J\varphi+\bar{J}\chi+\hat{J}\psi)}d\varphi\,d\chi\,d\psi. \quad (143)$$

Здесь J_i, $\bar{J}^\alpha$, $\hat{J}_\alpha$, K_i и L_α — *внешние источники* и «матричный элемент» (143) является функционалом от них; J_i и L_α— бозонные, а остальные — фермионные источники. Если функционал A заменить единицей, мы получим обобщение амплитуды «in — out»:

$$e^{iW\,[J,\ \bar{J},\ \hat{J},\ K,\ L,\ M]}\overset{\text{def}}{=}\langle\text{out}\,|\text{in}\rangle\overset{\text{def}}{=}\bar{N}\,(\det\gamma)^{1/2}\int e^{i\,(\tilde{S}+J\varphi+\bar{J}\chi+\hat{J}\psi)}d\varphi\,d\chi\,d\psi. \quad (144)$$

Это обобщение, аналогичное во многих отношениях статистической сумме в статистической механике, называется *производящим функционалом*, поскольку при разложении его в ряд по степеням источников, коэффициентами будут матричные элементы хронологических произведений полевых операторов. Коэффициент нулевого порядка есть исходная амплитуда (139).

На функционал $\tilde{S}$ можно смотреть как на обобщенный функционал действия. С помощью соотношений (97) и (98) легко показать, что он БРС-инвариантен. Допустим, что переменные φ, χ, ψ в интеграле (144) подвергнуты некоторому преобразованию БРС. Поскольку это немые переменные, интеграл остается неизменным. Поэтому

$$0=i\bar{N}\,(\det\gamma)^{1/2}\int\left(J_iQ^i_\alpha[\varphi]\,\psi^\alpha+\bar{J}^\alpha\gamma_{\alpha\beta}P^\beta[\varphi]-\frac{1}{2}\,\hat{J}_\alpha c^\alpha{}_{\beta\gamma}\psi^\beta\psi^\gamma\right)\times$$
$$\times e^{i\,(\tilde{S}+\hat{J}Q+\bar{J}\chi+\hat{J}\psi)}\,d\varphi\,d\chi\,d\psi. \quad (145)$$

С помощью равенства

$$0=\int\frac{\delta}{\delta\chi_\alpha}\{f\,[\varphi]e^{i\,(\tilde{S}+J\varphi+\bar{J}\chi+\hat{J}\psi)}\}\,d\varphi\,d\chi\,d\psi=$$
$$=i\int(F^\alpha{}_\beta[\varphi]\,\psi^\beta-\bar{J}^\alpha)\,f\,[\varphi]\,e^{i\,(\tilde{S}+J\varphi+\bar{J}\chi+\hat{J}\psi)}\,d\varphi\,d\chi\,d\psi, \quad (146)$$

где f — любая функция φ, этот результат можно представить в форме

$$0=i\bar{N}\,(\det\gamma)^{1/2}\int\left[J_i\frac{\delta\tilde{S}}{\delta K_i}+\frac{\delta\tilde{S}}{\partial M}-\hat{J}_\alpha\,\frac{\delta\tilde{S}}{\delta L_\alpha}\right]e^{i\,(\tilde{S}+J\varphi+\bar{J}\chi+\hat{J}\psi)}\,d\varphi\,d\chi\,d\psi=$$
$$=\left[J_i\frac{\delta}{\delta K_i}+\frac{\partial}{\partial M}-\hat{J}_\alpha\,\frac{\delta}{\delta L_\alpha}\right]e^{iW}, \quad (147)$$

или

$$J_i \frac{\delta W}{\delta K_i} + \frac{\partial W}{\partial M} - \hat{J}_\alpha \frac{\delta W}{\delta L_\alpha} = 0. \tag{148}$$

Уравнение (148) выражает важное свойство симметрии производящего функционала и прямо ведет к тождеству Уорда — Такахаши, которое рассматривается в разд. 5.8. Отметим, что производная по M является обычной частной производной, а не функциональной производной; M не несет никаких индексов и потому не зависит от пространства-времени.

5.7. МНОГОЧАСТИЧНЫЕ ФУНКЦИИ ГРИНА; ЭФФЕКТИВНОЕ ДЕЙСТВИЕ

В этом разделе существенным образом используется швингеровское среднее

$$\langle \mathbf{A} \rangle \overset{\text{def}}{=} \frac{\langle \text{out} \mid T(\mathbf{A}) \mid \text{in} \rangle}{\langle \text{out} \mid \text{in} \rangle}. \tag{149}$$

Здесь $\mathbf{A}$ — произвольный функционал операторов $\boldsymbol{\varphi}^i$, $\boldsymbol{\chi}_\alpha$, $\boldsymbol{\psi}^\alpha$, а числитель и знаменатель в правой части определены формулами (143) и (144). Если «in» и «out» — вакуумные состояния, а источники обращаются в нуль, это выражение превращается в швингеровское среднее (24). Удобно ввести

$$\varphi^i \overset{\text{def}}{=} \langle \boldsymbol{\varphi}^i \rangle, \quad \chi_\alpha \overset{\text{def}}{=} \langle \boldsymbol{\chi}_\alpha \rangle, \quad \psi^\alpha \overset{\text{def}}{=} \langle \boldsymbol{\psi}^\alpha \rangle. \tag{150}$$

Если источники K_i, L_α, M, $\bar{J}^\alpha$ и $\hat{J}_\alpha$ обращаются в нуль, средние χ_α и ψ^α обращаются в нуль и, кроме того, обращается в нуль J_i, то φ^i превращается в среднее, введенное в (25). Заметим, что хотя символы φ^i, χ_α и ψ^α выше были использованы для переменных интегрирования, никакой путаницы относительно их смысла на практике не возникнет.

Удобно также ввести коллективные обозначения: $\boldsymbol{\varphi}^A$ — для операторов $\boldsymbol{\varphi}^i$, $\boldsymbol{\chi}_\alpha$, $\boldsymbol{\psi}^\alpha$; φ^A — для их средних и J_A — для J_i, $\bar{J}^\alpha$, $\hat{J}_\alpha$. Пусть ΔJ_A — произвольные приращения источников. Тогда мы можем написать

$$\sum_{n=0}^{\infty} \frac{i^n}{n!} \Delta J_{A_n} \ldots \Delta J_{A_1} \langle \text{out} \mid T(\boldsymbol{\varphi}^{A_1} \ldots \boldsymbol{\varphi}^{A_n}) \mid \text{in} \rangle =$$
$$= \exp\left(\Delta J_A \frac{\delta}{\delta J_A}\right) \langle \text{out} \mid \text{in} \rangle = (e^{iW})_{J \to J + \Delta J} =$$
$$= \exp\left(iW + i\Delta J_A \varphi^A + i \sum_{n=2}^{\infty} \frac{1}{n!} \Delta J_{A_n} \ldots \Delta J_{A_1} G^{A_1 \ldots A_n}\right), \tag{151}$$

где

$$\varphi^A = \langle \boldsymbol{\varphi}^A \rangle = e^{-iW} \frac{\delta}{i \delta J_A} e^{iW} = \frac{\delta W}{\delta J_A}, \tag{152}$$

$$G^{A_1 \ldots A_n} \overset{\text{def}}{=} \frac{\delta}{\delta J_{A_1}} \ldots \frac{\delta}{\delta J_{A_n}} W. \tag{153}$$

Разделим обе части равенства (151) на e^{iW} и, сравнивая одинаковые степени ΔJ_A, получим бесконечную последовательность соотношений

$$\langle\varphi^A\varphi^B\rangle = \varphi^A\varphi^B - iG^{AB}, \tag{154}$$

$$\langle\varphi^A\varphi^B\varphi^C\rangle = \varphi^A\varphi^B\varphi^C - iP_3\varphi^A G^{BC} + (-i)^2 G^{ABC}, \tag{155}$$

и т. д., где P_N обозначает сумму по N различным перестановкам индексов, причем каждое слагаемое берется со знаком плюс или минус в зависимости от четности или нечетности перестановок фермионных индексов; G^{AB} — *одночастичный пропагатор*, а $G^{A_1 \ldots A_n}$, $n \geqslant 3$, — *многочастичные функции Грина* (в статистической механике они называются *корреляционными функциями*); они удовлетворяют граничным условиям, определяемым состояниями «in» и «out».

Любой функционал источников J_A можно также рассматривать как функционал средних φ^A. Из (152) и (153) видно, что одночастичный пропагатор является матрицей преобразования от одного набора переменных к другому:

$$G^{AB} = \frac{\delta\varphi^B}{\delta J_A}. \tag{156}$$

Этот факт можно использовать для установления важного соотношения между функционалом W и швингеровским средним от операторных полевых уравнений. Последние получаются из формального функционального тождества

$$0 = -ie^{-iW}\bar{N}(\det\gamma)^{1/2}\int e^{i(\tilde{S}+J\varphi+\bar{J}\chi+\hat{J}\psi)}\frac{\overleftarrow{\delta}}{\delta\varphi^A}\,d\varphi\,d\chi\,d\psi = \langle\tilde{\mathbf{S}}_{,A}\rangle + J_A. \tag{157}$$

Здесь $\tilde{\mathbf{S}}_{,A}$ — оператор, соответствующий функционалу $\tilde{S}_{,A}$, но ввиду присутствия фермионных полей мы должны теперь делать различие между правым и левым дифференцированиями. Когда все источники исчезают, в равенстве (157) индекс A становится равным i, и оно приводится к виду

$$0 = \langle\tilde{\mathbf{S}}_{,i}\rangle = \langle\mathbf{S}_{,i} + \mathbf{P}\gamma\mathbf{P}_{,i} - i(\ln\det\mathbf{F})_{,i}\rangle. \tag{158}$$

Это как раз уравнение (23) с добавлением упомянутых выше опущенных членов (содержащих $\mathbf{P}$ и $\det\mathbf{F}$).

Если мы продифференцируем (157) слева по J_B и воспользуемся соотношением (156), то получим

$$G^{BC}{}_{C,}\langle\tilde{\mathbf{S}}_{,A}\rangle = -\delta^B{}_A. \tag{159}$$

Здесь функциональная производная в скобках $\langle\ldots\rangle$ берется по полевому оператору φ^A, а функциональная производная вне скобок берется по усредненному полю φ^C. Легко видеть, что ${}_{C,}\langle\tilde{S}_{,A}\rangle$ есть оператор, одночастичным пропагатором для которого явля-

ется функция Грина. Можно показать, что в силу своих граничных условий G^{BC} является одновременно и левой, как в (159), и правой функцией Грина. Она обладает следующей симметрией:

$$G^{AB} = (-1)^{AB} G^{BA}, \tag{160}$$

откуда следует

$$_{A,}\langle \tilde{S}_{,B} \rangle = (-1)^{A+B+AB}\, _{B,}\langle \tilde{S}_{,A} \rangle, \tag{161}$$

где подразумевается, что индексы в показателе (—1) принимают значения +1 или —1 в зависимости от того, бозонные они или фермионные. Но (161) представляет собой как раз условие существования функционала $\tilde{\Gamma}[\varphi, \chi, \psi, K, L, M]$, такого, что

$$\tilde{\Gamma}_{,A} = \langle \tilde{S}_{,A} \rangle. \tag{162}$$

Функционал Г известен как *эффективное действие*. Он удовлетворяет уравнениям

$$\tilde{\Gamma}_{,A} = -J_A, \tag{163}$$

$$_{A,}\tilde{\Gamma}_{,C} G^{CB} = -\delta_A^B \tag{164}$$

и связан с функционалом W преобразованием Лежандра

$$W = \tilde{\Gamma} + J_A \varphi^A. \tag{165}$$

Это последнее соотношение можно проверить дифференцированием по J_B и использованием уравнения (163) в виде $_{A,}\tilde{\Gamma} = (-1)^A J_A$. Тогда имеем

$$\frac{\delta W}{\delta J_B} = \frac{\delta \varphi^A}{\delta J_B} [_{A,}\tilde{\Gamma} + (-1)^A J_A] + \varphi^B = \varphi^B, \tag{166}$$

а это есть как раз соотношение (152). Поскольку функционал $\tilde{\Gamma}$ определен с точностью до произвольной постоянной интегрирования, равенство (165) можно рассматривать как фиксирующее эту постоянную.

Функционал $\tilde{\Gamma}$ называют также *производящим функционалом собственных вершин*. Это название основано на его связи с многочастичными функциями Грина. Дифференцируя (164), можно связать функциональные производные одночастичного пропагатора с производными $\tilde{\Gamma}$. Эти соотношения приводят, например, к такому равенству:

$$G^{ABC} = \frac{\delta}{\delta J_A} G^{BC} = G^{AD}{}_{D,}G^{BC} = \\ = (-1)^{(B+C)D+(C+D)E+(D+E)F} G^{AD} G^{BE} G^{CF}{}_{DEF,}\tilde{\Gamma}. \tag{167}$$

Если пропагаторы изображать линиями, а третьи и более высокие производные $\tilde{\Gamma}$ — вершинами, легко увидеть, что каждое новое дифференцирование по источнику добавляет всеми возможными

способами новую линию в предыдущие диаграммы. Таким образом, каждая функция Грина данного порядка представима как сумма всевозможных древесных диаграмм этого порядка. Из того, что (при существовании асимптотических областей) S-матрицу можно выразить через хронологические произведения, появляющиеся в (151), и из того, что эти произведения выражаются через функции Грина (154), (155), следует, что при использовании $\tilde{\Gamma}$ для построения S-матрицы нужны только древесные диаграммы. Никаких замкнутых петель не появляется. Вершины, порожденные функционалом $\tilde{\Gamma}$, являются *собственными* вершинами; в них уже включены все квантовые поправки. Замечая, что аналогичные древесные диаграммы появляются в классической теории возмущений, но с заменой $\tilde{\Gamma}$ на $\tilde{S}$, можно показать, что $\tilde{\Gamma}$ описывает динамику когерентных полей большой амплитуды с квантовыми поправками. Это должно быть справедливо и в случаях, когда не существует асимптотических областей и нет S-матрицы.

5.8. ТОЖДЕСТВО УОРДА — ТАКАХАШИ

Вернемся теперь к использованию символов φ^i, χ_α, ψ^α, J_i, $\bar{J}^\alpha$, $\hat{J}_\alpha$ и перепишем (165) в более явной форме:

$$W[J, \bar{J}, \hat{J}, K, L, M] = \tilde{\Gamma}[\varphi, \chi, \psi, K, L, M] + J_i\varphi^i + \bar{J}^\alpha\chi_\alpha + \hat{J}_\alpha\psi^\alpha. \tag{168}$$

Средние φ^i, χ_α, ψ^α зависят от всех шести источников, но, поскольку K_i, L_α, M не участвуют в преобразовании Лежандра, можно показать, что

$$\frac{\delta W}{\delta K_i} = \frac{\delta \tilde{\Gamma}}{\delta K_i}, \quad \frac{\delta W}{\delta L_\alpha} = \frac{\delta \tilde{\Gamma}}{\delta L_\alpha}, \quad \frac{\partial W}{\partial M} = \frac{\partial \tilde{\Gamma}}{\partial M}, \tag{169}$$

где производные в правых частях относятся только к явной зависимости $\tilde{\Gamma}$ от K_i, L_α, M. Этот результат в сочетании с уравнением (163) в виде ${}_{A,}\tilde{\Gamma} = (-1)^A J_A$ позволяет переписать (148) в виде

$$-\frac{\delta \tilde{\Gamma}}{\delta \varphi^i}\frac{\delta \tilde{\Gamma}}{\delta K_i} + \frac{\partial \tilde{\Gamma}}{\partial M} - \frac{\delta \tilde{\Gamma}}{\delta \psi^\alpha}\frac{\delta \tilde{\Gamma}}{\delta L_\alpha} = 0. \tag{170}$$

Это и есть *тождество Уорда — Такахаши.*

Тождество Уорда — Такахаши имеет важную связь со структурой $\tilde{\Gamma}$. Тот факт, что из него следует существование определенного рода симметрии, присущей $\tilde{\Gamma}$, становится очевидным, если заметить, что в силу своей БРС-инвариантности $\tilde{S}$ также удовлетворяет тождеству Уорда — Такахаши. К сожалению, вывести из соотношения (170) симметрию $\tilde{\Gamma}$ — намного более трудная задача. В принципе можно было бы поступить следующим образом. Предположим, что $\tilde{\Gamma}$ можно разложить в ряд по степеням χ_α, ψ^α,

K_i и L_α. Такое предположение априори не является обращением к теории возмущений, поскольку это разложение должно быть произведено *после* взятия функционального интеграла (144). Оно основано на серьезной уверенности в том, что (144) ведет себя гладко (по крайней мере после соответствующих перенормировок) при устремлении к нулю K_i, L_α, M, $\bar{J}^\alpha$ и $\hat{J}_\alpha$ (и, следовательно, χ_α и ψ^α).

При определении того, какие типы членов могут появиться в разложении, полезно ввести понятие «ду́хова числа». Если приписывать «ду́ховы числа»: 1 — полю ψ^α и источнику $\bar{J}^\alpha$, 0 — полю φ^i и источнику J_i, —1 — полю χ_α и источникам $\hat{J}_\alpha$, K_i, M и —2 — источнику L_α, то легко увидеть, что подынтегральное выражение в (144) и сам интеграл имеют полное «ду́ховое число» 0. Следовательно, W и $\hat{\Gamma}$ имеют полное «ду́ховое число» 0, и все члены в разложении также обладают этим свойством. Кроме того, разложение не может содержать члены старше первой степени по M, поскольку M — антикоммутирующая постоянная.

Если подставить это разложение в равенство (170) и собрать вместе члены одинаковых степеней, то получится бесконечная последовательность дополнительных тождеств Уорда — Такахаши, связывающих коэффициенты, зависящие от φ. К сожалению, по-видимому, нет никакого легкого способа получить из этих тождеств какие-либо простые выводы относительно ситуации в целом. До сих пор тождество (170) применялось лишь к перенормируемым моделям по теории возмущений порядок за порядком. В этих случаях тождество сослужило большую службу как в практических деталях осуществления ренормализационной программы, так и при демонстрации того, что теория действительно перенормируема во всех порядках и при этом сохраняется унитарность.

Тождество Уорда — Такахаши должно играть столь же важную роль в квантовой гравитации. В частности, ожидается, что оно приведет к следующему результату, который, как обнаружилось, выполняется в перенормируемых теориях: когда все источники обращаются в нуль, уравнение (163) (с $A=i$) должно эффективно сводиться к уравнению

$$\Gamma_{,i}[\varphi]=0, \tag{171}$$

где функционал Γ калибровочно-инвариантен:

$$\Gamma_{,i}Q^i{}_\alpha\equiv 0. \tag{172}$$

В квантовой гравитации (171) есть как раз уравнение (26), причем Γ есть *приведенное эффективное действие*. Однако доказательство того, что все на самом деле именно так и обстоит, остается пока программой на будущее.

6. ЗАКЛЮЧЕНИЕ

Любой обзор по квантовой гравитации всегда заканчиваешь с острым чувством того, как много еще остается непознанного, как много еще необходимо сделать. Наши представления о структуре группы диффеоморфизмов и о типах глобальных трудностей, которые могут встретиться, носят пока весьма неполный характер. Эти трудности могут оказаться относительно безобидными, но могут и привести к огромным осложнениям. Как пример того, насколько следует быть осторожным, рассмотрим введение канонических координат в качестве способа получить (в принципе) глобально выполняющиеся калибровочные условия. Канонические координаты основаны на однопараметрических абелевых подгруппах. В отличие от положения в большинстве непрерывных групп однопараметрические абелевы подгруппы группы диффеоморфизмов не заполняют окрестность единицы (см. [41]). Имеются определенные C^{∞}-диффеоморфизмы, сколь угодно близкие к единице, которые не могут быть достигнуты из единицы по экспоненте инфинитезимального преобразования. Весьма вероятно, что это безобидный факт. Структура любой ковариантной теории и фактически самой группы диффеоморфизмов почти несомненно определяется уже теми диффеоморфизмами, которые получаются из единицы по экспоненте. И мы должны приветствовать все, что снижает «размеры» группы.

Возможен вопрос: почему мы вообще столь пространно обсуждаем эту группу? Почему не перейти прямо к инвариантам? Ведь именно в них содержится вся физика. Ответ таков: было бы прекрасно, если бы это можно было сделать. Но это не так легко. Мы не можем, например, иметь дело с локальными полями (тензорами, спинорами), не вводя калибровочной группы, и пока еще никто не указал, как сформулировать принцип действия без локальных полей.

Эти замечания делаются вовсе не для того, чтобы бросить тень на будущее квантовой гравитации. Несмотря на трудности, все то, что достигнуто до сих пор, дает право «болеть» за нее. Полученные результаты столь красивы, что нельзя в них не верить. Открытие Хокингом квантовых черных дыр уже показывает, что эта теория, как никакая до нее, соединяет теорию относительности, теорию квантов и статистическую механику в одно гармоническое целое. Вызывает чувство удовлетворения тот замечательный факт, что именно этим областям принадлежит неоценимое наследие Эйнштейна.

Ирония, однако, состоит в том, что сам Эйнштейн, вероятно, воспринял бы все наши усилия с усмешкой, как попытку путаников соединить несоединимое. Для него квантовая теория была не тем, что должно быть со временем наложено на общую теорию относительности, а скорее тем, что должно быть вытеснено рано

или поздно некоторой нелинейной детерминистической теорией, построенной, возможно, по тем же принципам, что и сама общая теория относительности. И если иметь в виду его обостренное восприятие глубочайших вопросов физики, кто знает, не прав ли он?

Несомненно, квантовая гравитация затрагивает эти глубокие вопросы, и очень жаль, что Эйнштейн не дожил до того, чтобы увидеть по крайней мере первые ростки своего посева. Прежде чем закончить настоящий обзор, осталось упомянуть об одном из самых глубоких вопросов. Большинство эффектов, которые обсуждались в первой части обзора (эффект Казимира, влияние топологии и кривизны на $\langle T^{\mu\nu} \rangle$, тепловое излучение), крайне малы и становятся доминирующими лишь в таких экстремальных условиях, какие трудно себе представить: для этого необходимо увеличение плотности вещества в 10^{80} раз. Эти условия соответствуют последним стадиям гравитационного коллапса при значениях массы от массы звезды до массы вселенной в целом. Тогда и в макроскопических масштабах полностью властвует квантовая теория.

Если вселенная в целом была в какой-то момент или будет снова квантовым объектом, что же такое тогда объективность, служащая той почвой, на которой мы стоим как ученые? Если мы являемся частью волновой функции, как мы можем наблюдать действительность так, как мы это делаем? На эти вопросы имеется только один ответ, у которого есть хоть какой-нибудь шанс на успех, без того чтобы изменить известную нам структуру квантовой механики,— это ответ, предложенный Эвереттом [32]. За годы, прошедшие со времени скромного появления первой статьи Эверетта, его концепция привлекала все возрастающее число видных сторонников. Эта концепция сохраняет как основу экспериментально подтвержденные, абсолютные статистические законы квантовой механики, но возвращается к детерминизму в глобальном масштабе, вводя всеобщую (grand) волновую функцию и рассматривая ее как такое описание действительности, которое само определяет ход мысли, если понимать его буквально.

Когда пытаются определить динамическое поведение вселенной в ее первые моменты, вычисляя эффективное действие [44], или ищут механизм изотропизации ранней вселенной, вычисляя рождение частиц [51, 72], в конечном счете всегда сталкиваются с проблемами интерпретации квантовой механики и с концепцией Эверетта. Именно общая теория относительности, соединенная в новом синтезе с квантовой теорией, настоятельно выдвигает эти вопросы на первый план.

ЛИТЕРАТУРА

1. *Abers E. S., Lee B. W.*, Phys. Rep., **9C**, 1 (1973).
2. *Atiyah M. F., Singer I. M.*, Ann. Math., **87**, 546 (1968).
3. *Atiyah M. F., Bott R., Patodi V. K.*, Invent Math., **19**, 279 (1973).
4. *Becchi C., Rouet A., Stora R.*, Commun. Math. Phys., **42**, 127 (1975).

5. *Bekenstein J. D.*, Phys. Rev., **D7**, 2333 (1973).
6. *Березин Ф. А.* Метод вторичного квантования.— М.: Наука, 1965.
7. *Bohr N.*, *Rosenfeld L.*, Kgl. Danske Videnskab. Selskab. Mat.-fys. Med., **12**, No. 8 (1933).
8. *Brown L. S.*, *Cassidy J. P.*, Phys. Rev., **D15**, 2810 (1977).
9. *Brown L. S.*, *Cassidy J. P.*, Phys. Rev., **D16**, 1712 (1977).
10. *Bunch T. S.*, *Davies P. C. W.*, Proc. R. Soc. Lond., **A356**, 569 (1977).
11. *Candelas P.*, *Deutsch D.*, Proc. R. Soc. Lond., **A354**, 79 (1977).
12. *Capper D. M.*, *Duff M. J.*, Nuovo Cimento, **23A**, 173 (1974).
13. *Casimir H. B. G.*, Proc. Kon. Ned. Akad. Wetenschap., **51**, 793 (1940).
14. *Christiensen S. M.*, *Fulling S. A.*, Phys. Rev., **D15**, 2088 (1977). [Имеется перевод: Черные дыры.— М.: Мир, с. 261, 1978.]
15. *Davies P. C. W.*, *Fulling S. A.*, *Unruh W. G.*, Phys. Rev., **D13**, 2720 (1976).
16. *Delbourgo R.*, *Ramon Medrano M.*, Nucl. Phys., **B110**, 467 (1976).
17. *Delbourgo R.*, *Salam A.*, Phys. Lett., **40B**, 381 (1972).
18. *Deser S.*, *Duff M. J.*, *Isham C. J.*, Nucl. Phys., **B111**, 45 (1976).
19. *DeWitt B. S.*, In: Gravitation: An Introduction to Current Research, ed. L. Witten. Wiley, New York, 1962.
20. *DeWitt B. S.*, Phys. Rev. Lett., **13**, 114 (1964).
21. *DeWitt B. S.*, Dynamical Theory of Groups and Fields. Gordon and Breach, New York, 1965.
22. *DeWitt B. S.*, Phys. Rev., **160**, 1113 (1967).
23. *DeWitt B. S.*, Phys. Rev., **162**, 1195, 1239 (1967).
24. *DeWitt B. S.*, In: Relativity: Proceedings of the Relativity Conference in the Midwest, eds. M. Carmeli, S. I. Fickler, L. Witten. Plenum Press, New York, 1970.
25. *DeWitt B. S.* In: Magic Without Magic, ed. J. Klauder. W. H. Freeman, San Francisco, 1972.
26. *Dixon J. D.*, Phil. thesis. University of Oxford, 1975.
27. *Dowker J. S.*, *Critchley R.*, Phys. Rev., **D13** 3224 (1976).
28. *Dowker J. S.*, *Kennedy G.*, J. Phys., **11A**, 895 (1978).
29. *Duff M. J.*, Phys. Rev., **D7**, 2317 (1973).
30. *Eguchi T.*, *Freund P. G. O.*, Phys. Rev. Lett., **37**, 1251 (1976).
31. *Einstein A.*, Mein Weltbild. Querido Verlag: Amsterdam, 1933.
32. *Everett H.*, Rev. Mod. Phys., **29**, 454. Перепечатано вместе с несколькими другими статьями по этому же вопросу в сб. The Many-Worlds Interpretation of Quantum Mechanics, eds. B. S. DeWitt, R. N. Graham. Princeton University Press, 1973.
33. *Фаддеев Л. Д.* ТМФ **1**, 3 (1969).
34. *Fadde'ev L. D.*, In: Methods in Field Theory, eds. R. Balian, J. Zinn-Justin (1975 Les Houches Lectures). North-Holland, Amsterdam.
35. *Fadde'ev L. D.*, *Popov V. N.*, Phys. Lett., **25B**, 29 (1967).
36. *Fetter A. L.*, *Walecka J. D.*, Quantum Theory of Many-Particle Systems. McGraw-Hill, New York, 1971.
37. *Feynman R. P.*, Rev. Mod. Phys., **20**, 267 (1948).
38. *Feynman R. P.*, Phys. Rev., **80**, 440 (1950).
39. *Feynman R. P.*, Acta Phys. Polon., **24**, 697 (1963); см. также в Proceedings of the 1962 Warsaw Conference on the Theory of Gravitation, PWN-Editions Scientifiques de Pologne, Warszawa, 1964.
40. *Fischer A. E.*, In: Relativity: Proceedings of the Relativity Conference in the Midwest, eds. M. Carmeli, S. I. Fickler, I. Witten Plenum Press, New York, 1970.
41. *Freifeld C.*, In: Battelle Rencontres (1967 Lectures in Mathematics and Physics) eds. C. M. DeWitt, J. A. Wheeler. Benjamin, New York, 1968.
42. *Gilkey P. B.*, J. Differ. Geom., **10**, 601 (1975).
43. *Грибов В. Н.* В сб. Физика элементарных частиц (Материалы XII зимней школы ЛИЯФ), Л., 1977.
44. *Hartle J. B.*, Phys. Rev. Lett., **39**, 1373 (1977).

45. *Hawking S. W.*, Phys. Rev. Lett., **26**, 1344 (1971).
46. *Hawking S. W.*, Nature, **248**, 30 (1974).
47. *Hawking S. W.*, Phys. Lett., **60A**, 81 (1977).
48. *Hawking S. W.*, Commun. Math. Phys., **55**, 133 (1977).
49. *Hawking S. W.*, Phys. Rev., **D18**, 1447 (1978).
50. *Higgs P. Ж.*, Phys. Rev. Lett., **12**, 132 (1964); 13, 508. [См. также Phys. Rev., **145**, 1156 (1966).]
51. *Hu B. L.*, *Parker L.*, Phys. Rev., **D17**, 933 (1978).
52. *Isaacson R. A.*, Phys. Rev., **166**, 1263, 1272 (1968).
53. *Isham C. J.*, Proc. R. Soc. Lond., **A362**, 383 (1978).
54. *Isham C. J.*, *Strathdee J.*, *Salam A.*, Phys. Rev., **D3**, 1805 (1971).
55. *Isham C. J.*, *Strathdee J.*, *Salam A.*, Phys. Rev., **D5**, 2584 (1972).
56. *Jackiw R.*, Rev. Mod. Phys., **49**, 681 (1977).
57. *Jackiw R.*, *Rebbi C.*, Phys. Rev., **D16**, 1052 (1977).
58. *Kimura T.*, Prog. Theor. Phys., **42**, 1191 (1969).
59. *Kostant B.*, In: Differential Geometrical Methods in Mathematical Physics (Proceedings of the July 1—4, 1975 Symposium in Bonn), eds. K. Bleuler, A. Reetz. Lecture Notes in Mathematics No. 570. Springer, Berlin, 1977.
60. *Lee B. W.*, In: Methods in Field Theory (1975 Les Houches Lectures), eds. R. Balian, J. Zinn-Justin. North-Holland, Amsterdam, 1976.
61. *McKean H. P.*, *Singer I. M.*, J. Differ. Geom., **1**, 43 (1967).
62. *Press W.*, *Teukolsky S.*, Nature, **238**, 211 (1972).
63. *Rindler W.*, Am. J. Phys., **34**, 1174, (1966).
64. *Schwinger J.*, Proc. Nat. Acad. Sci. USA, **37**, 452, 455 (1951).
65. *Stelle K. S.*, Phys. Rev., **D16**, 953 (1977).
66. *'t Hooft G.*, Nucl. Phys., **B35**, 167 (1971).
67. *'t Hooft G.*, Phys. Rev. Lett., **37**, 8 (1976).
68. *'t Hooft G.*, Phys. Rev., **D14**, 3422 (1976).
69. *Unruh W. G.*, Phys. Rev., **D14**, 870 (1976).
70. *Utiyama R.*, Phys. Rev., **101**, 1597 (1956). [Имеется перевод: Элементарные частицы и компенсирующие поля.— М.: Мир, 1964.]
71. *Yang C. N.*, *Mills R.*, Phys. Rev., **96**, 191 (1954). [Имеется перевод: Элементарные частицы и компенсирующие поля.— М.: Мир, 1964.]
72. *Зельдович Я. Б.*, Письма ЖЭТФ, **12**, 443 (1970).
73. *Зельдович Я. Б.*, Письма ЖЭТФ, **14**, 270 (1971).
74. *Зельдович Я. Б.*, ЖЭТФ, **62**, 2076 (1972).

VII. ИНТЕГРАЛЫ ПО ТРАЕКТОРИЯМ В ПРИЛОЖЕНИИ К КВАНТОВОЙ ГРАВИТАЦИИ

С. Хокинг [1])

1. ВВЕДЕНИЕ

Классическая общая теория относительности — вполне законченная теория. Она дает не только уравнения, которым подчиняется гравитационное поле, но и уравнения движения тел под действием этого поля. Однако есть два обстоятельства, которые не позволяют ей дать удовлетворительное во всех отношениях описание наблюдаемой вселенной. Во-первых, она трактует гравитационное поле как чисто классическое, в то время как все другие поля, по-видимому, квантованные. Во-вторых, из ряда теорем (см. [31]) следует, что общая теория относительности неизбежно ведет к сингулярностям пространства-времени. Как предсказывает эта теория, сингулярности имеют место в начале нынешнего расширения вселенной («большой взрыв») и при коллапсе звезд с образованием черных дыр. Вблизи этих сингулярностей классическая общая теория относительности полностью теряет силу или становится неполной, поскольку она не может предсказать, что именно должно поступать из сингулярности (иными словами, она не обеспечивает для уравнений поля краевые условия в сингулярных точках). По обеим упомянутым причинам было бы желательно построить квантовую теорию гравитации. Какого-либо рецепта вывода такой теории из классической общей теории относительности нет. Приходится, прибегая к интуиции и общим соображениям, строить теорию, которая была бы полна, самосогласованна и находилась бы в согласии с классической теорией относительности для макроскопических тел и при малой кривизне пространства-времени. Нужно сразу сказать, что пока не существует теории, которая удовлетворяла бы указанным трем критериям, особенно первому и второму. Но получен ряд частных результатов, которые столь убедительны, что трудно себе представить, чтобы они не были частью окончательной полной картины. Эти результаты относятся к связи между черными дырами и термодинамикой. (Такая связь обсуждается, например, в обзорных статьях Картера [48] и Гиббонса [49].) В настоящей статье будет показано, как это соответствие между гравитацией и термодинамикой проявляется при квантовании самого гравитационного поля.

Имеются три основных подхода к квантованию гравитации.

---

[1]) *S. W. Hawking*, Department of Applied Mathematics and Theoretical Physics, Silver Street, Cambridge, UK.

I. ОПЕРАТОРНЫЙ ПОДХОД

В этом подходе в классических уравнениях Эйнштейна метрика заменяется оператором в некотором гильбертовом пространстве со значениями в пространстве обобщенных функций. Однако эта процедура не представляется особенно подходящей в применении к теории, подобной теории гравитации, где полевые уравнения не полиномиальны. Довольно трудно придать смысл даже произведению полевых операторов в одной и той же точке пространства-времени, не говоря уж о неполиномиальной функции вроде обращения метрики или квадратного корня из ее определителя.

II. КАНОНИЧЕСКИЙ ПОДХОД

В этом подходе вводится семейство пространственноподобных поверхностей, которое используется для построения гамильтониана и канонических коммутационных соотношений в один и тот же момент времени. Ряд авторов отдает предпочтение этому подходу, поскольку он, по-видимому, применим к сильным гравитационным полям и, как представляется, обеспечивает унитарность. Но расщепление на три пространственных измерения и одно время, по-видимому, противоречит самому духу теории относительности. Более того, оно ограничивает топологию пространства-времени произведением действительной прямой на некоторое трехмерное многообразие, в то время как надо ожидать, что квантовая гравитация будет допускать все возможные топологии пространства-времени, включая и те, которые не являются прямыми произведениями. Как раз такие топологии, по-видимому, и приводят к наиболее интересным эффектам. Имеется также проблема смысла одновременных коммутационных соотношений. Последние корректно определены для материальных полей при фиксированной геометрии пространства-времени, но какой смысл говорить, что две точки пространственноподобно разделены, если геометрия квантована и подчиняется принципу неопределенности?

По этим причинам я предпочитаю третий подход.

III. ПОДХОД, ОСНОВАННЫЙ НА ИНТЕГРИРОВАНИИ ПО ТРАЕКТОРИЯМ

Этот подход связан с рядом трудностей и нерешенных проблем, но он, по-видимому, подает наибольшие надежды. Отправной точкой для этого подхода служит идея Фейнмана, согласно которой амплитуду

$$\langle g_2, \phi_2, S_2 | g_1, \phi_1, S_1 \rangle$$

перехода от состояния с метрикой g_1 и материальными полями ϕ_1 на поверхности S_1 к состоянию с метрикой g_2 и материальными полями ϕ_2 на поверхности S_2 можно представить как сумму по всем конфигурациям полей g и ϕ, которые принимают данные зна-

чения на поверхностях S_1 и S_2 (рис. 1). Точнее

$$\langle g_2, \phi_2, S_2 | g_1, \phi_1, S_1 \rangle = \int D[g, \phi] \exp(iI[g, \phi]),$$

где $D[g, \phi]$ — некоторая мера в пространстве всех конфигураций полей g и ϕ, $I[g, \phi]$ — действие для этих полей и интеграл берется по всем полям, которые имеют заданные значения S_1 и S_2.

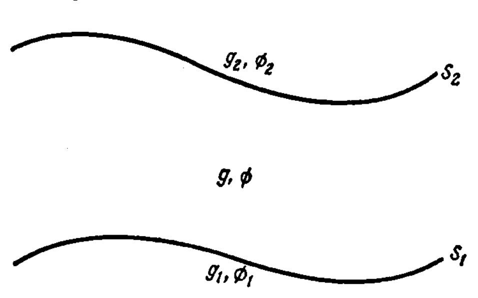


Рис. 1. Амплитуда $\langle g_2, \phi_2, S_2 | g_1, \phi_1, S_1 \rangle$ перехода от метрики g_1 и материальных полей ϕ_1 на поверхности S_1 к метрике g_2 и материальным полям ϕ_2 на поверхности S_2 определяется интегралом по траектории по всем полям g, ϕ, которые имеют данные значения на S_1 и S_2.

Здесь неявно предполагается одно из двух: либо поверхности S_1 и S_2 и область между ними компактны («замкнутая» вселенная), либо гравитационное и материальные поля спадают некоторым подходящим образом на пространственной бесконечности (асимптотически плоское пространство). Чтобы последнему случаю придать более точный смысл, нужно соединить поверхности S_1 и S_2 времениподобной трубкой большого радиуса так, чтобы эта граница и область внутри ее были компактны, как в случае замкнутой вселенной. Как будет показано в следующем разделе, поверхность на бесконечности играет важную роль из-за присутствия поверхностного члена в гравитационном действии.

Не все компоненты метрик g_1 и g_2 на границе имеют физический смысл, поскольку компонентам $g^{ab}n_b$ можно придать произвольные значения диффеоморфизмами или калибровочными преобразованиями, которые сдвигают точки внутри области M, но оставляют неизменной границу ∂M. Таким образом, нужно только задать индуцированную трехмерную метрику h на ∂M и лишь с точностью до диффеоморфизмов, отображающих границу на себя.

В последующих разделах будет показано, как подход, основанный на таком интеграле по траекториям, может быть применен к квантованию гравитации и как он ведет к понятиям температуры черной дыры и квантовомеханической внутренней энтропии.

2. ДЕЙСТВИЕ

В общей теории относительности действие обычно берется в виде

$$I=\frac{1}{16\pi G}\int(R-2\Lambda)(-g)^{1/2}d^4x+\int L_m(-g)^{1/2}d^4x, \quad (1)$$

где R — скалярная кривизна, Λ — космологическая постоянная, g — детерминант метрики и L_m — лагранжиан материальных полей. Используемые единицы таковы, что $c=\hbar=k=1$. G — ньютоновская постоянная; иногда я буду пользоваться единицами, в которых и $G=1$. При вариациях метрики, которые обращаются в нуль вместе со своими нормальными производными на границе ∂M компактной области M, это действие стационарно, если и только если метрика удовлетворяет уравнениям Эйнштейна

$$R_{ab}-\frac{1}{2}g_{ab}R+\Lambda g_{ab}=8\pi G T_{ab}, \quad (2)$$

где $T^{ab}=\frac{1}{2}(-g)^{-1/2}(\delta L_m/\delta g_{ab})$ — тензор энергии-импульса материальных полей. Однако действие не будет экстремальным, если допускаются вариации метрики, которые сами обращаются в нуль на границе, но их нормальные производные не обращаются в нуль. Причина состоит в том, что скалярная кривизна R содержит члены, которые линейны по вторым производным метрики. Интегрированием по частям вариация этих членов может быть превращена в интеграл по границе, который содержит нормальные производные вариации метрики на границе. Для того чтобы устранить этот поверхностный интеграл и получить, таким образом, действие, стационарное для решений уравнений Эйнштейна при всех вариациях метрики, исчезающих на границе, нужно добавить к действию член вида [18]

$$\frac{1}{8\pi G}\int K(\pm h)^{1/2}d^3x+C, \quad (3)$$

где K — след второй фундаментальной формы границы, h — индуцированная метрика на границе; знаки плюс или минус выбираются в зависимости от того, пространственноподобна или времениподобна граница, а C — член, который зависит только от метрики на границе ∂M, но не от значений метрики во внутренних точках. Необходимость добавления к действию поверхностного члена (3) в подходе с интегрированием по траекториям можно увидеть, рассмотрев ситуацию, изображенную на рис. 3, где рассматривается переход от метрики h_1 на поверхности S_1 к метрике h_2 на поверхности S_2 и затем к метрике h_3 на более поздней поверхности S_3. Потребуем, чтобы амплитуда перехода от начального состояния к конечному получалась суммированием всех состояний на промежуточной поверхности S_2, т. е.

$$\langle h_3, S_3|h_1S_1\rangle=\sum_{h_2}\langle h_2, S_2|h_1,S_1\rangle\langle h_3, S_3|h_2,S_2\rangle. \quad (4)$$

Это справедливо, если и только если

$$I[g_1+g_2]=I[g_1]+I[g_2], \quad (5)$$

где g_1 — метрика между S_1 и S_2, g_2— метрика между S_2 и S_3, а $[g_1+g_2]$ — метрика в областях между S_1 и S_3, полученных соединением двух прежних областей. Поскольку нормальная производная g_1 на S_2, вообще говоря, не равна нормальной производ-

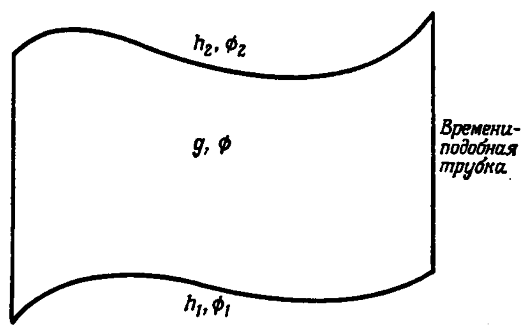


Рис. 2. На граничной поверхности нужно задать только индуцированную метрику h.

В асимптотически плоском случае начальную и конечную поверхности следует соединить времениподобной трубкой большого радиуса, чтобы область, над которой производится интегрирование по траекториям, была компактной.

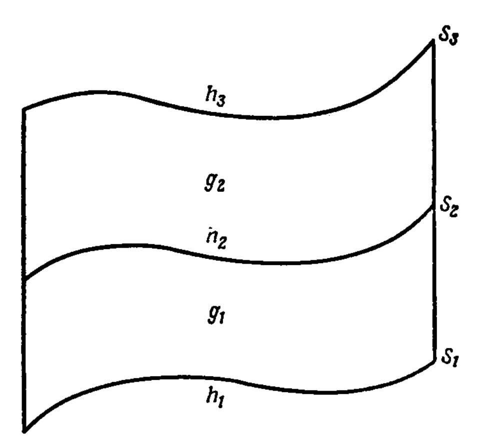


Рис. 3. Амплитуда перехода от метрики h_1 на поверхности S_1 к метрике h_3 на поверхности S_3 должна быть суммой амплитуд перехода во все метрики h_2 на промежуточной поверхности S_2.

Это условие будет выполняться, только когда действие содержит поверхностный член.

ной g_2 на S_2, метрика $[g_1+g_2]$ будет иметь в тензоре Риччи дельта-функцию с множителем $2(K^1_{ab}-K^2_{ab})$, где K^1_{ab} и K^2_{ab} — вторые фундаментальные формы поверхности S_2 при метриках g_1 и g_2 соответственно; они определены по отношению к нормали, на-

правленной в будущее. Это означает, что соотношение (5) справедливо, если и только если действие является суммой (1) и (3), т. е.

$$I=\frac{1}{16\pi G}\int (R-2\Lambda)(-g)^{1/2}d^4x+\int L_m(-g)^{1/2}d^4x+ \\ +\frac{1}{8\pi G}\int K(\pm h)^{1/2}d^3x+C. \quad (6)$$

Появление члена C в выражении для действия довольно неприятно. Его можно было бы просто включить в перенормировку меры $D[g, \phi]$. Но в случае асимптотически-плоской метрики естественно взять его в таком виде, чтобы вклад от времениподобной трубки при больших радиусах был равен нулю, когда g совпадает с метрикой плоского пространства: $g=\eta$. Тогда

$$C=-\frac{1}{8\pi G}\int K^0(\pm h)^{1/2}d^3x, \quad (7)$$

где K^0 — вторая фундаментальная форма границы, вложенной в плоское пространство. Это не вполне удовлетворительный рецепт, так как, вообще говоря, метрика границы h не может быть вложена в плоское пространство. Однако в асимптотически-плоской ситуации можно предположить, что граница становится асимптотически-вложимой при увеличении радиуса. Я подозреваю, что в конечном счете следует отбросить все граничные поверхности и иметь дело только с замкнутыми пространственно-временными многообразиями. Однако при нынешнем уровне развития теории очень удобно использовать некомпактные асимптотически-плоские метрики и вычислять действие, используя границу при большом радиусе.

Метрика, асимптотически-плоская в трех пространственных измерениях, но не во времени, может быть записана в виде

$$ds^2=-(1-2M_t r^{-1})dt^2+(1+2M_s r^{-1})dr^2+ \\ +r^2(d\theta^2+\sin^2\theta\, d\phi^2)+O(r^{-2}). \quad (8)$$

Если эта метрика удовлетворяет вакуумным уравнениям Эйнштейна ($\Lambda=0$) вблизи бесконечности, то $M_t=M_s$; но в интеграле по траекториям рассматриваются все асимптотически-плоские метрики независимо от того, удовлетворяют они уравнениям Эйнштейна или нет. В такой метрике удобно выбрать границу ∂M в виде прямого произведения оси t на сферу радиуса r_0. Площадь такой границы ∂M равна

$$\int(-h)^{1/2}d^3x=4\pi r_0^2\int(1-2M_t r_0^{-1}+O(r_0^{-2}))\,dt. \quad (9)$$

Интеграл следа второй фундаментальной формы ∂M равен

$$\int K(-h)^{1/2}d^3x=\frac{\partial}{\partial n}\int(-h)^{1/2}d^3x, \quad (10)$$

где $\partial/\partial n$ обозначает производную в направлении сдвига каждой точки ∂M вдоль единичной нормали. Таким образом,

$$\int K(-h)^{1/2} d^3x = \int (8\pi r_0 - 4\pi M_t - 8\pi M_s + O(r_0^{-2}))\, dt. \quad (11)$$

Для плоской метрики η имеем $K^0 = 2r_0^{-1}$. Отсюда

$$\frac{1}{8\pi G}\int (K - K^0)(-h)^{1/2} d^3x = \frac{1}{2G}\int (M_t - 2M_s)\, dt. \quad (12)$$

В частности, для решения уравнения Эйнштейна с массой M (при измерении с бесконечности) $M_s = M_t = M$, и поверхностный член имеет вид

$$-\frac{M}{2G}\int dt + O(r_0^{-1}). \quad (13)$$

3. КОМПЛЕКСНОЕ ПРОСТРАНСТВО-ВРЕМЯ

Для действительной лоренцевой метрики g (т. е. метрики с сигнатурой $-+++$) и действительных материальных полей ϕ действие $I[g, \phi]$ будет действительным; следовательно, интеграл по траекториям будет осциллировать и не будет сходиться. С этим связана такая трудность: чтобы найти конфигурацию поля, которая отвечает экстремальному действию между данными начальной и конечной поверхностями, нужно решить гиперболическое уравнение с начальными и конечными граничными условиями. Но это некорректная задача: у нее может вообще не быть решений или их может оказаться бесконечное число, а если и будет какое-то решение, то оно не будет зависеть гладко от граничных значений.

В обычной квантовой теории поля в плоском пространстве-времени эту трудность преодолевают путем поворота оси времени на 90° по часовой стрелке в комплексной плоскости, т. е. t заменяется на $i\tau$. При этом в интеграл по объему от действия I вводится множитель $-i$. Например, для скалярного поля массы m лагранжиан имеет вид

$$L = -\frac{1}{2}\phi_{,a}\phi_{,b}g^{ab} - \frac{1}{2}m^2\phi^2. \quad (14)$$

Поэтому функциональный интеграл

$$Z = \int D[\phi] \exp(iI[\phi]) \quad (15)$$

принимает вид

$$Z = \int D[\phi] \exp(-\hat{I}[\phi]), \quad (16)$$

где $\hat{I} = -iI$ — «евклидово» действие, причем $\hat{I} \geqslant 0$ для полей ϕ, которые действительны в евклидовом пространстве, определяемом действительными τ, x, y, z. Таким образом, интервал по всем таким конфигурациям поля будет экспоненциально подавлен и, следо-

вательно, должен сходиться. Далее, при замене t мнимой координатой τ лоренцева метрика η^{ab} (сигнатура $-+++$) заменяется на евклидову (сигнатура $++++$). При этом задача о нахождении экстремума действия становится корректной задачей о решении эллиптического уравнения с данными краевыми значениями.

Идея состоит в том, чтобы выполнить все интегрирования по траекториям в евклидовом секторе (τ, x, y, z действительны) и затем аналитически продолжить против часовой стрелки в комплексной t-плоскости в лоренцев сектор (сектор Минковского), где t, x, y, z действительны. В качестве примера рассмотрим величину

$$Z[J]=\int D[\phi]\exp\left(-\left(\frac{1}{2}\phi A\phi+J\phi\right)\right)dx\,dy\,dz\,d\tau, \tag{17}$$

где A — дифференциальный оператор второго порядка $-\Box+m^2$, $\Box$ — четырехмерный лапласиан и $J(x)$ — заданное поле-источник, которое убывает на больших евклидовых расстояниях. Интеграл по траекториям берется по всем полям ϕ, которые убывают на больших евклидовых расстояниях. Символически $Z[J]$ можно записать в виде

$$Z[J]=\exp\left(\frac{1}{2}JA^{-1}J\right)\int D[\phi]\exp\left(-\frac{1}{2}(\phi-A^{-1}J)\,A(\phi-A^{-1}J)\right), \tag{18}$$

где $A^{-1}(x_1, x_2)$ — единственный обратный к A оператор (функция Грина), который убывает на больших евклидовых расстояниях:

$$A^{-1}J(x)=\int A^{-1}(x, x')\,J(x')\,d^4x', \tag{19}$$

$$JA^{-1}J=\iint J(x)\,A^{-1}(x, x')\,J(x')\,d^4x\,d^4x'. \tag{20}$$

Мера $D[\phi]$ инвариантна относительно трансляций $\phi\to\phi-A^{-1}J$. Таким образом,

$$Z[J]=\exp(^1/_2JA^{-1}J)\,Z[0]. \tag{21}$$

Тогда мы можем определить евклидов пропагатор (двухточечную функцию корреляций)

$$\langle 0|\phi(x_2)\phi(x_1)|0\rangle=\left.\frac{\delta^2\ln Z}{\delta J(x_1)\,\delta J(x_2)}\right|_{J=0}=A^{-1}(x_2, x_1). \tag{22}$$

Фейнмановский пропагатор получается аналитическим продолжением $A^{-1}(x_2, x_1)$ против часовой стрелки в комплексной t_2-t_1-плоскости.

Следует подчеркнуть, что использование евклидова сектора позволяет задать вакуумное состояние требованием, чтобы поля ϕ убывали при больших положительных и отрицательных значениях мнимого времени τ. Операция упорядочения по времени, применяемая при определении фейнмановского пропагатора, автоматически воссоздается направлением аналитического продолжения из евк-

лидова пространства, поскольку если $\operatorname{Re}(t_2-t_1)>0$, то $\langle 0|\phi(x_2), \phi(x_1)|0\rangle$ является голоморфной функцией в нижней t_2-t_1-полуплоскости, т. е. положительно-частотной функцией (положительно-частотная функция — это функция, голоморфная в нижней t-полуплоскости и убывающая при больших мнимых t).

Другим важным для последующего применением евклидова сектора является построение канонического ансамбля для поля ϕ. Амплитуда распространения от конфигурации ϕ_1 на некоторой поверхности в момент t_1 к конфигурации ϕ_2 на некоторой поверхности в момент t_2 дается функциональным интегралом

$$\langle\phi_2,\ t_2|\phi_1,\ t_1\rangle=\int D[\phi]\exp(iI[\phi]). \tag{23}$$

Используя картину Шредингера, можно также записать эту амплитуду в виде

$$\langle\phi_2|\exp(-iH(t_2-t_1))|\phi_1\rangle.$$

Положим $t_2-t_1=-i\beta$, $\phi_2=\phi_1$ и просуммируем по полному ортонормированному базису конфигураций ϕ_n. Получим статистическую сумму

$$Z=\sum\exp(-\beta E_n) \tag{24}$$

поля ϕ при температуре $T=\beta^{-1}$, где E_n — энергия состояния ϕ_n. Однако, согласно (23), можно также представить Z как евклидов интеграл по траекториям

$$Z=\int D[\phi]\exp(-i\hat{I}[\phi]), \tag{25}$$

где интеграл берется по всем полям ϕ, которые действительны в евклидовом секторе и периодичны по мнимой координате времени τ с периодом β. Как и прежде, можно ввести источник J и получить функцию Грина функциональным дифференцированием $Z[J]$ по J в двух различных точках. Эта производная будет представлять собой двухточечную корреляционную функцию, или пропагатор, для поля ϕ, но в данном случае не в вакуумном состоянии, а в каноническом ансамбле при температуре $T=\beta^{-1}$. В пределе, когда период β стремится к бесконечности, этот тепловой пропагатор стремится к нормальному вакуумному фейнмановскому пропагатору.

Представляется разумным применить аналогичные идеи комплексификации к гравитационному полю, т. е. к метрике. Допустим, например, что рассматривается амплитуда перехода от метрики h_1 на поверхности S_1 к метрике h_2 на поверхности S_2, где поверхности S_1 и S_2 асимптотически-плоские и разделены на бесконечности интервалом времени t. Как объяснялось в разд. 1, следует соединить S_1 и S_2 времениподобной трубкой длины t с очень большим радиусом. Тогда можно повернуть этот временно́й интервал в комплексную плоскость путем введения мнимой координаты

времени $\tau=it$. Индуцированная на времениподобной трубке метрика будет тогда положительно-определенной, так что мы будем иметь дело с интегралом по траекториям над областью M, на границе которой индуцированная метрика h всюду положительно определена. Тогда интеграл по траекториям можно рассматривать как интеграл по всем положительно-определенным метрикам g, которые индуцируют на ∂M данную положительно-определенную метрику h. При том же выборе направления поворота в комплексную плоскость, как в плоской евклидовой теории, множитель $(-g)^{1/2}$, появляющийся в элементе объема, превращается в $-i(g)^{1/2}$, так что евклидово действие $\hat{I}=-iI$ становится равным

$$\hat{I}=-\frac{1}{16\pi G}\int(R-2\Lambda)(g)^{1/2}d^4x-\frac{1}{8\pi G}\int(K-K^0)(h)^{1/2}d^3x-\\ -\int L_m(g)^{1/2}d^4x. \quad (26)$$

Проблема, возникающая из-за того, что гравитационная часть этого евклидова действия не является положительно-определенной, рассматривается в разд. 4.

Состояние системы определяется выбором граничных условий для метрик, по которым производится интегрирование. Например, представляется разумным ожидать, что вакуумное состояние будет соответствовать интегрированию по всем метрикам, которые асимптотически евклидовы, т. е. вне некоторого компактного множества они приближаются к плоской евклидовой метрике на R^4. Внутри этого компактного множества кривизна может быть большой, а топология — отличной от топологии R^4.

В качестве примера можно рассмотреть канонический ансамбль для гравитационных полей, содержащихся в сферической полости радиуса r_0 при температуре T, путем интегрирования по траекториям по всем метрикам, которые заполняют область внутри границы, состоящей из времениподобной трубки радиуса r_0, которая периодически отождествляется с периодом $\beta=T^{-1}$ в направлении мнимого времени.

При комплексификации пространственно-временного многообразия величины, которые комплексны в действительном лоренцевом секторе, и комплексно-сопряженные им величины приходится рассматривать как независимые величины. Например, заряженное скалярное поле ϕ в действительном лоренцевом пространстве-времени можно описывать комплексным полем ϕ и сопряженным ему полем $\bar{\phi}$. При переходе к комплексному пространству-времени необходимо аналитически продолжать $\bar{\phi}$ как некоторое новое поле $\tilde{\phi}$, которое не зависит от ϕ. То же самое относится к спинорам. В действительном лоренцевом пространстве-времени мы имеем нештрихованные спиноры λ_A, которые преобразуются по $SL(2, C)$, и штрихованные спиноры $\mu_{A'}$, преобразующиеся по ком-

плексно-сопряженной группе $\overline{SL(2, C)}$. Комплексно-сопряженный нештрихованный спинор есть штрихованный спинор, и наоборот. При переходе к комплексному пространству-времени штрихованные и нештрихованные спиноры становятся совершенно независимыми друг от друга и преобразуются соответственно по независимым группам $SL(2, C)$ и $\widetilde{SL}(2, C)$. Если делается аналитическое продолжение в сектор, в котором метрика положительно-определена, и накладывается ограничение, чтобы спиноры лежали только в этом секторе, то штрихованные и нештрихованные спиноры по-прежнему независимы, но группами преобразований становятся соответственно $SU(2)$ и $\widetilde{SU}(2)$. Например, в лоренцевой метрике тензор Вейля может быть представлен в виде

$$C_{AA'BB'CC'DD'} = \psi_{ABCD}\varepsilon_{A'B'}\varepsilon_{C'D'} + \bar{\psi}_{A'B'C'D'}\varepsilon_{AB}\varepsilon_{CD}. \tag{27}$$

При комплексификации $\bar{\psi}_{A'B'C'D'}$ заменяется независимым полем $\tilde{\psi}_{A'B'C'D'}$. В частности, может быть метрика, в которой $\psi_{ABCD} \neq 0$, но $\tilde{\psi}_{A'B'C'D'} = 0$. Такую метрику называют конформно самодуальной, и для нее имеет место равенство

$$C_{abcd} = {}^*C_{abcd} = \frac{1}{2}\varepsilon_{abef}C^{ef}{}_{cd}. \tag{28}$$

Метрика называется самодуальной, если

$$R_{abcd} = {}^*R_{abcd},$$

откуда следует

$$R_{ab} = 0, \quad C_{abcd} = {}^*C_{abcd}. \tag{29}$$

Комплексифицированное пространственно-временно́е многообразие M с комплексной самодуальной или конформно-самодуальной метрикой g_{ab} может допускать сектор, в котором метрика действительна и положительно-определена («евклидов» сектор), но не допускает лоренцева сектора, т. е. сектора, в котором метрика действительна и обладает сигнатурой $-+++$.

4. НЕОПРЕДЕЛЕННОСТЬ ГРАВИТАЦИОННОГО ДЕЙСТВИЯ

Евклидово действие для скалярных полей или полей Янга — Миллса положительно-определено. Это означает, что интеграл по траекториям в евклидовом секторе сходится и вклад в него дают только такие конфигурации, которые убывают на больших евклидовых расстояниях, поскольку иначе действие становится бесконечным. Для фермионных полей действие не является положительно-определенным. Однако они рассматриваются как антикоммутирующие величины [1], так что интеграл по траекториям по ним сходится. Евклидово гравитационное действие, напротив, не является положительно-определенным даже при действительной положи-

тельно-определенной метрике. Причина состоит в том, что, хотя гравитационные волны несут положительную энергию, гравитационная потенциальная энергия отрицательна из-за того, что гравитация есть притяжение. Несмотря на это, в классической общей теории относительности полная энергия или масса любого асимптотически-плоского гравитационного поля, измеренные с бесконечности, всегда неотрицательны. Это утверждение известно как *гипотеза о положительности энергии* [3, 15]. Всякий раз, когда гравитационная потенциальная энергия становится слишком большой, по-видимому, происходит следующее: образуется горизонт событий, и область большой гравитационной связи испытывает гравитационный коллапс, оставляя после себя черную дыру положительной массы. Таким образом, можно было бы ожидать, что черные дыры играют определенную роль в проблеме неопределенности гравитационного действия в квантовой теории, и есть указания на то, что это в самом деле так.

Чтобы увидеть, как это действие может принимать произвольные отрицательные значения, рассмотрим конформное преобразование $\tilde{g}_{ab}=\Omega^2 g_{ab}$, где Ω — положительная функция, равная единице на границе ∂M; имеем

$$\tilde{R}=\Omega^{-2}R-6\Omega^{-3}\,\Box\,\Omega, \tag{30}$$

$$\tilde{K}=\Omega^{-1}K+3\Omega^{-2}\Omega_{;a}n^a, \tag{31}$$

где n — единичная внешняя нормаль к границе ∂M. Итак,

$$\hat{I}[\tilde{g}]=-\frac{1}{16\pi G}\int\limits_M(\Omega^2 R+6\Omega_{;a}\Omega_{;b}g^{ab}-2\Lambda\Omega^4)(g)^{1/2}d^4x-$$

$$-\frac{1}{8\pi G}\int\limits_{\partial M}\Omega^2(K-K^0)(h)^{1/2}d^3x. \tag{32}$$

Отсюда видно, что $\hat{I}$ может принимать любое отрицательное значение, если выбрать быстро изменяющийся конформный множитель Ω.

При изучении этой проблемы, по-видимому, желательно разбить интегрирование по всем метрикам на интегрирование по конформным множителям с последующим интегрированием по конформно-эквивалентным классам метрик. Я буду рассматривать по отдельности случай, когда космологическая постоянная Λ равна нулю, но пространственно-временнáя область имеет границу ∂M, и случай, когда $\Lambda\neq 0$, а область компактна без границы.

В первом случае функциональный интеграл по траекториям по конформному множителю определяется конформно-инвариантным скалярным волновым оператором $A=-\Box+{}^1/_6R$. Пусть $\{\lambda_n,\phi_n\}$ — собственные значения и собственные функции оператора A с граничными условиями Дирихле, т. е.

$$A\phi_n=\lambda_n\phi_n,\ \phi_n=0 \text{ на } \partial M.$$

Если $\lambda_i=0$, то $\Omega^{-1}\phi_i$ есть собственная функция с нулевым собственным значением для метрики $\tilde{g}_{ab}=\Omega^2 g_{ab}$. Ненулевые собственные значения и соответствующие собственные функции не обладают таким простым поведением при конформных преобразованиях. Но они будут изменяться непрерывно при гладких вариациях конформного множителя, остающегося всюду положительным. Поскольку нулевое собственное значение конформно-инвариантно, число отрицательных собственных значений (которое конечно) остается неизменным при конформных преобразованиях с положительными всюду Ω.

Пусть $\Omega=1+y$, где $y=0$ на ∂M. Тогда

$$\hat{I}[\tilde{g}]=-\frac{6}{16\pi G}\int(yAy+2Ry)(g)^{1/2}d^4x+\hat{I}[g]=$$

$$=-\frac{6}{16\pi G}\int\{(y-A^{-1}R)A(y-A^{-1}R)\}(g)^{1/2}d^4x=$$

$$=\frac{6}{16\pi G}RA^{-1}R+\hat{I}[g]=\frac{6}{16\pi G}RA^{-1}R+\hat{I}[g]-\frac{6}{16\pi G}\int\gamma A\gamma(g)^{1/2}d^4x, \tag{33}$$

где $\gamma=(y-A^{-1}R)$.

Таким образом, можно написать

$$\hat{I}[\tilde{g}]=I^1+I^2,$$

где I^1 — первый и второй члены в правой части выражения (33), а I^2 — третий член.

I^1 зависит только от классов конформной эквивалентности метрики g, тогда как I^2 зависит от конформного множителя. Следовательно, можно ввести величину X, которая будет интегралом по траекториям от $\exp(-I^2)$ по всем конформным множителям в одном классе конформно-эквивалентных метрик.

Если оператор A не имеет ни отрицательных, ни нулевых собственных значений, в частности если g есть решение уравнений Эйнштейна, то оператор A^{-1} всюду определен и метрика $g'_{ab}=(1+A^{-1}R)^2 g_{ab}$ будет регулярной метрикой, причем $R'=0$ всюду. В этом случае $I^1=\hat{I}[g']$, а $\hat{I}[g']$ в свою очередь определяется поверхностным интегралом от K по границе. Представляется правдоподобной следующая *гипотеза о положительности действия*: любая асимптотически-евклидова положительно-определенная метрика с $R=0$ имеет положительное или равное нулю действие [23]. Между этой гипотезой и гипотезой о положительности энергии в классической лоренцевой общей теории относительности есть тесная связь. Последняя утверждает, что измеренные с бесконечности масса и энергия любого лоренцева, асимптотически-плоского решения уравнений Эйнштейна положительны или равны нулю, если это решение развивается с некоторой несингулярной начальной поверхности, причем масса будет равна нулю, если и только если метрика тождественно-плоская. Хотя какого-либо полного

доказательства не существует, все же гипотеза о положительности энергии подтверждается в ряде ограниченных случаев или при определенных допущениях [2, 3, 15, 34], и в нее все верят. Если бы она выполнялась и в пятимерной классической общей теории относительности (сигнатура $-++++$), то отсюда бы следовала гипотеза о положительности действия, поскольку четырехмерную асимптотически-евклидову метрику с $R=0$ можно было бы взять в качестве симметричных по времени начальных данных для пятимерного решения и масса такого решения была бы равна действию для пятимерной метрики. Пейдж [38] получил ряд результатов, свидетельствующих в пользу гипотезы о положительности действия. Однако он показал также, что эта гипотеза не выполняется для метрик типа решения Шварцшильда, асимптотически-плоских в пространственных направлениях, но не в направлении евклидова времени. Смысл этого будет разъяснен ниже.

Пусть g_0 есть решение полевых уравнений. Если I^1 возрастает при всех возмущениях g_0, которые не являются чисто конформными преобразованиями, интеграл по конформным классам будет сходиться. Если есть какое-либо неконформное возмущение δg метрики g_0, при котором I^1 убывает, то для того, чтобы функциональный интеграл сходился, мы должны проинтегрировать по всем метрикам вида $g_0+i\delta g$. При этом в Z будет внесен множитель i для каждой моды неконформных возмущений, при которой I^1 убывает. Этот вопрос мы обсудим в следующем разделе. При метриках, которые далеки от решения полевых уравнений, у оператора A могут оказаться нулевые или отрицательные собственные значения. Когда собственное значение проходит через нуль, A^{-1} становится неопределенным, а I^1— бесконечным. Когда имеются отрицательные собственные значения, но нет нулевых, A^{-1} и I^1 определены, но конформный множитель $\Omega=1+A^{-1}R$, который преобразует метрику g к метрике g' с $R'=0$, будет проходить через нуль, и, таким образом, метрика g' будет сингулярной. Это весьма похоже на то, что происходит с трехмерными метриками на симметричных по времени начальных поверхностях [2]. Если h — трехмерная положительно-определенная метрика на начальной поверхности, мы можем совершить конформное преобразование $\tilde{h}=\Omega^4 h$, чтобы получить метрику с $\tilde{R}=0$, которая будет удовлетворять уравнениям связи. Если трехмерный конформно-инвариантный оператор $B=-\Delta+R/8$ не имеет ни нулевых, ни отрицательных собственных значений (что действительно имеет место для метрик h, достаточно близких к плоскому пространству), конформный множитель Ω будет необходимо конечным и всюду положительным. Но если рассматривать последовательность метрик h, для которой одно из собственных значений оператора B переходит через нуль и становится отрицательным, то соответствующий множитель Ω будет сначала расходиться, потом снова

станет конечным, но будет проходить через нуль, и метрика $\tilde{h}$ будет сингулярной. Объясняется это тем, что метрика h имеет область со столь большой отрицательной гравитационной энергией связи, что она сама отделяет себя от остальной вселенной, образуя горизонт событий. Чтобы описывать подобного рода ситуации, мы должны использовать начальные поверхности с различными топологиями.

По-видимому, нечто аналогичное может происходить в четырехмерном случае. В некотором смысле можно считать, что метрика g, при которой оператор A имеет отрицательные собственные значения, содержит области, которые отделяют сами себя от остального пространства-времени, так как в них слишком много кривизны. Их вклад можно тогда учесть, переходя к многообразиям с другими топологиями. Во всяком случае, метрики, при которых A имеет отрицательные собственные значения, являются в некотором смысле далекими от решений полевых уравнений, и в следующем разделе мы увидим, что в действительности мы можем ограничиться вычислением интегралов по траекториям только по метрикам, близким к решениям полевых уравнений.

Оператор A появляется в I^2 со знаком минус. Это значит, что интеграл по траекториям по конформным множителям сходится в окрестности решения полевых уравнений, в частности в окрестности плоского пространства, лишь при условии, что γ взята чисто мнимой. Поэтому рецепт получения сходящегося интеграла по траекториям состоит в разделении пространства всех метрик на конформно-эквивалентные классы. В каждом классе эквивалентности выделяется метрика g', для которой $R'=0$. Далее проводится интегрирование по всем метрикам $\tilde{g}=\Omega^2 g'$, где Ω имеет вид $1+i\xi$. Затем проводится интегрирование по классам конформной эквивалентности вблизи решений полевых уравнений, причем неконформные возмущения должны быть чисто мнимыми для мод, при которых I^1 убывает.

Похожая ситуация имеет место для компактных многообразий с Λ-членом. В этом случае в действии нет ни поверхностного члена, ни требования $\Omega=1$ на границе. Если $\tilde{g}=\Omega^2 g$, то

$$\hat{I}[\tilde{g}] = -\frac{6}{16\pi G}\int (\Omega^2 R + 6\Omega_{;a}\Omega_{;b}g^{ab} - 2\Lambda\Omega^4)(g)^{1/2}\, d^4x. \tag{34}$$

Таким образом, квантовая гравитация с Λ-членом на компактном многообразии представляет собой нечто вроде усреднения теории $\lambda\phi^4$ по всем фоновым метрикам. Но в отличие от обычной теории $\lambda\phi^4$ кинетический член $(\nabla\Omega)^2$ присутствует в действии со знаком минус. Это значит, что по конформным множителям нужно интегрировать в комплексном направлении, как и в предыдущем случае.

Пространство всех положительно-определенных метрик g на многообразии M можно снова разбить на классы конформной эквивалентности. В каждом классе эквивалентности действие будет

иметь один экстремум при нулевой метрике, для которой $\Omega=0$. В общем случае будет еще и другой экстремум при метрике g', для которой $R'=4\Lambda$, хотя в некоторых случаях конформное преобразование $g'=\Omega^2 g$, где g — положительно-определенная метрика, может потребовать комплексного множителя Ω. Положив $\tilde{g}=(1+y)^2 g'$, получим

$$\hat{I}[\tilde{g}]=-\frac{\Lambda V}{8\pi G}-\frac{6}{16\pi G}\int(6y_{;a}y_{;b}g^{ab}-8y^2\Lambda-8y^3\Lambda-2y^4\Lambda)(g')^{1/2}d^4x, \qquad (35)$$

где $V=\int(g')^{1/2}d^4x$.

Если $\Lambda<0$ и мы пренебрегаем членами третьей и четвертой степеней по y, то приходим к сходящемуся интегралу по траекториям при интегрировании по чисто мнимым y аналогично тому, как это происходило в предыдущем случае. Поэтому представляется разумным следовать такому рецепту вычисления интегралов по траекториям с Λ-членом: в каждом классе конформной эквивалентности выделяется метрика g', для которой $R'=4\Lambda$, и затем производится интегрирование по конформным множителям вида $\Omega=1+i\xi$ в окрестности g'.

Если $\Lambda>0$, оператор $-6\Box-8\Lambda$, который действует на квадратичные по ξ члены, обладает по крайней мере одним отрицательным собственным значением $\xi=\text{const}$. На самом деле это, по-видимому, единственное отрицательное собственное значение. Его смысл обсуждается в разд. 10.

5. ПРИБЛИЖЕНИЕ СТАЦИОНАРНОЙ ФАЗЫ

Ожидается, что главный вклад в интеграл по траекториям вносят метрики и поля, которые близки к метрике g_0 и полям ϕ_0, определяющим экстремум действия, т. е. к решениям классических уравнений. Пренебрегая на время вопросом сходимости, можно разложить действие в ряд Тейлора в окрестности фоновых полей g_0, ϕ_0:

$$\hat{I}[g,\phi]=\hat{I}[g_0,\phi_0]+I_2[\bar{g},\bar{\phi}]+\text{члены высших порядков}, \qquad (36)$$

где

$$g_{ab}=g_{0\,ab}+\bar{g}_{ab}, \qquad \phi=\phi_0+\bar{\phi}$$

и $I_2[\bar{g},\bar{\phi}]$ — член, квадратичный по возмущениям $\bar{g}$ и $\bar{\phi}$. Если игнорируются члены высших порядков, интеграл по траекториям имеет вид

$$\ln Z=-\hat{I}[g_0,\phi_0]+\ln\int D[\bar{g},\bar{\phi}]\exp(-I_2[\bar{g},\bar{\phi}]). \qquad (37)$$

Это приближение называют по-разному: приближением стационарной фазы, ВКБ-приближением или однопетлевым приближе-

нием. Первый член в правой части (37) можно рассматривать как вклад фонового поля в $\ln Z$. Мы обсудим его в разд. 7 и 8. Второй член называется однопетлевым членом и представляет собой вклад квантовых флуктуаций относительно фоновых полей. Остающаяся часть настоящего раздела посвящена изложению способа вычисления этого вклада. Для простоты будет рассмотрен только случай, когда все фоновые материальные ϕ_0 поля равны нулю. Тогда квадратичный член $I_2[\bar{g}, \bar{\phi}]$ можно представить как $I_2[\bar{g}]+I_2[\phi]$ и

$$\ln Z = -\hat{I}[g_0]+\ln\int D[\phi]\exp(-I_2[\phi])+\ln\int D[\bar{g}]\exp(-I_2[\bar{g}]). \tag{38}$$

Сначала рассмотрим однопетлевой член для материальных полей, т. е. второй член в правой части равенства (38). $I_2[\phi]$ можно записать в виде

$$I_2[\phi]=\frac{1}{2}\int \phi A\phi\,(g_0)^{1/2}d^4x, \tag{39}$$

где A — дифференциальный оператор, зависящий от фоновой метрики g_0. В случае бозонных полей, который будет рассмотрен первым, A есть дифференциальный оператор второго порядка. Пусть $\{\lambda_n, \phi_n\}$ — собственные значения и собственные функции оператора A, причем в случае, когда есть граничная поверхность, $\phi_n=0$ на ∂M. Собственные функции ϕ_n можно нормировать:

$$\int \phi_n\phi_m\cdot(g_0)^{1/2}d^4x=\delta_{nm}. \tag{40}$$

Произвольное поле ϕ, которое обращается в нуль на ∂M, можно выразить в виде линейной комбинации этих собственных функций:

$$\phi=\sum_n y_n\phi_n. \tag{41}$$

Подобным же образом меру в пространстве всех полей ϕ можно записать в виде

$$D[\phi]=\prod_n \mu\, dy_n, \tag{42}$$

где μ — нормировочный множитель размерности массы или обратной длины. Тогда однопетлевой член для материи можно представить как

$$Z_\phi=\int D[\phi]\exp(-I_2[\phi])=\prod_n\int \mu\, dy_n\exp\left(-\frac{1}{2}\lambda_n y_n^2\right)=$$
$$=\prod_n(2\pi\mu^2\lambda_n^{-1})^{1/2}=\left(\det\left(\frac{1}{2}\pi^{-1}\mu^{-2}A\right)\right)^{-1/2}. \tag{43}$$

В случае комплексного поля ϕ типа заряженного скалярного поля

нужно рассматривать ϕ и аналитическое продолжение $\tilde{\phi}$ комплексно-сопряженного к ϕ поля как независимые поля. Тогда квадратичный член имеет вид

$$I_2[\phi, \tilde{\phi}] = \frac{1}{2}\int \tilde{\phi} A \phi (g_0)^{1/2} d^4x. \quad (44)$$

Если есть фоновое электромагнитное поле, то оператор A не будет самосопряженным. Можно записать $\tilde{\phi}$ через собственные функции сопряженного оператора $A^\dagger$:

$$\tilde{\phi} = \sum_n \tilde{y}_n \tilde{\phi}_n. \quad (45)$$

Тогда мера будет иметь вид

$$D[\phi, \tilde{\phi}] = \prod_n \mu^2 dy_n d\tilde{y}_n. \quad (46)$$

Поскольку мы интегрируем по y_n и $\tilde{y}_n$ независимо, получаем

$$Z_\phi = \left(\det\left(\frac{1}{2}\pi^{-1}\mu^{-2}A\right)\right)^{-1}. \quad (47)$$

Чтобы интегралы по траекториям применить к фермионам, спинор ψ и его независимый сопряженный спинор $\tilde{\psi}$ нужно считать антикоммутирующими грассмановыми числами [1]. Для грассмановой переменной x имеют место следующие (формальные) правила интегрирования:

$$\int dx = 0, \quad \int x\,dx = 1. \quad (48)$$

Этого достаточно, чтобы определить все интегралы, поскольку x^2 и более высокие степени x равны нулю по свойству антикоммутации. Отметим, что из (48) следует $dy = a^{-1}dx$, если $y = ax$, где a — действительная постоянная.

Эти правила можно применить к вычислению интеграла по траектории по ферми-полям ψ и $\tilde{\psi}$. Оператором A в этом случае будет обычный оператор Дирака первого порядка. Если $\exp(-I_2)$ разложить в степенной ряд, то из-за свойства антикоммутации останется лишь член, линейный по A. Интегрирование по $d\psi$ и $d\tilde{\psi}$ дает

$$Z_\psi = \det\left(\frac{1}{2}\mu^{-2}A\right). \quad (49)$$

Таким образом, однопетлевые члены для ферми-полей пропорциональны детерминанту их оператора, тогда как для бозонов они обратно пропорциональны детерминантам.

Можно получить следующее асимптотическое разложение для числа собственных значений $N(\lambda)$ оператора A со значениями

меньше λ:

$$N(\lambda) \sim \frac{1}{2} B_0\lambda^2 + B_1\lambda + B_2 + O(\lambda^{-1}), \tag{50}$$

где B_0, B_1, B_2 — коэффициенты Адамара — Минакшисундарама — Де Витта (АМДВ), о которых упоминает Гиббонс [49]. Их можно представить в виде

$$B_n = \int b_n (g_0)^{1/2} d^4x,$$

где b_n — скалярные полиномы по метрике, кривизне и ее ковариантным производным [25]. В случае скалярного волнового оператора $A = -\Box + \xi R + m^2$ они равны

$$b_0 = \frac{1}{16\pi^2}, \tag{51}$$

$$b_1 = \frac{1}{16\pi^2}((1/6 - \xi) R - m^2), \tag{52}$$

$$b_2 = \frac{1}{2880\pi^2}(R^{abcd}R_{abcd} - R_{ab}R^{ab} + (6-30\xi)\Box R + \frac{5}{2}(6\xi - 1)^2 R^2) + \\ + 30m^2(1 - 6\xi) R + 90m^4). \tag{53}$$

Когда имеется граничная поверхность ∂M, отсюда возникают дополнительные вклады в (50), включающие члены порядка $\lambda^{1/2}$. По-видимому, это должно быть дополнительным основанием для попыток отбросить граничные поверхности и работать просто с замкнутыми многообразиями.

Из разложения (50) видно, что детерминант A, т. е. произведение его собственных значений, сильно расходится. Чтобы получить конечный ответ, мы должны регуляризовать этот детерминант, поделив его на произведение собственных значений, соответствующих первым двум членам в правой части (50) (и членам порядка $\lambda^{1/2}$, если они есть). Для этого существуют разные способы: размерная регуляризация [44], расщепление точек [8], регуляризация Паули — Вилларса [47] и регуляризация через дзета-функцию [9, 29, 30]. Последний метод представляется наиболее подходящим для регуляризации детерминантов операторов на фоне искривленного пространства. Мы обсудим его в следующем разделе.

Как для фермионных, так и бозонных операторов член B_0 равен $(nV/16\pi^2)$, где V — объем многообразия в фоновой метрике, а n — число спиновых состояний поля. Поэтому, если число бозонных и фермионных спиновых состояний одинаково, ведущие расходимости в Z, порожденные членами B_0, взаимно сократятся в фермионных и бозонных детерминантах без регуляризации. Если к тому же члены B_1 сокращаются или равны нулю (что имеет место для безмассовых конформно-инвариантных полей), то остающиеся главные расходимости взаимно сокращаются между бозонами и фермионами. Такая ситуация имеет место в теориях с суперсим-

метрией типа супергравитации [5, 14] или расширенной супергравитации [12]. Это может служить достаточным основанием для того, чтобы серьезно относиться к этим теориям, в частности к связи материальных полей с гравитацией.

Независимо от того, взаимно сокращаются или устраняются регуляризацией сингулярности, возникающие от B_0 и B_1, остающийся коэффициент B_2 будет, вообще говоря, отличен от нуля, даже в супергравитации, если топология пространства-времени нетривиальна [39]. Это значит, что выражение для Z будет содержать конечное число (не обязательно целое) нескомпенсированных собственных значений. Поскольку эти собственные значения имеют размерность квадрата обратной длины, чтобы получить безразмерный результат для Z, каждое из них нужно поделить на μ^2, где μ — нормирующая постоянная, или регуляризующая масса. Таким образом, Z зависит от μ. Для перенормируемых теорий, таких, как $\lambda\phi^4$, квантовая электродинамика или теория Янга — Миллса в плоском пространстве, коэффициент B_2 пропорционален действию данного поля. Это означает, что зависимость от μ можно включить в эффективную константу взаимодействия $g(\mu)$, зависящую от масштаба, в котором она измеряется. Если $g(\mu)\to 0$ при $\mu\to\infty$, т. е. для очень коротких расстояний или высоких энергий, то говорят, что теория — асимптотически-свободная.

Но в искривленном пространстве B_2 содержит члены, квадратичные по тензору кривизны фонового пространства. Поэтому без предположения о том, что гравитационное действие содержит члены, квадратичные по кривизне (что, по-видимому, ведет к множеству проблем, включая отрицательность энергии, уравнения четвертого порядка и отсутствие ньютоновского предела [42, 43]), нельзя устранить зависимость от μ. Поэтому говорят, что гравитация неперенормируема: при регуляризации теории появляются новые параметры.

Если бы мы попытались регуляризовать члены более высокого порядка в ряде Тейлора относительно фоновой метрики, то пришлось бы ввести бесконечную последовательность параметров регуляризации, значения которых не фиксировались бы теорией. Однако в разд. 9 утверждается, что эти члены высшего порядка не имеют физического смысла и следует рассматривать лишь однопетлевые квадратичные члены. В отличие от теорий Янга — Миллса или $\lambda\phi^4$ в теории гравитации есть естественный масштаб длины — планковская масса. Поэтому представляется разумным взять нормирующий множитель μ для однопетлевого члена кратным планковской массе.

6. РЕГУЛЯРИЗАЦИЯ ЧЕРЕЗ ДЗЕТА-ФУНКЦИЮ

Чтобы регуляризовать детерминант оператора A с собственными значениями и собственными функциями $\{\lambda_n, \phi_n\}$, построим из

собственных значений обобщенную дзета-функцию:

$$\zeta_A(s) = \sum \lambda_n^{-s}. \qquad (54)$$

Из разложения (50) можно видеть, что ζ будет сходиться при $\mathrm{Re}\, s > 2$. Можно аналитически продолжить ζ до мероморфной функции s с полюсами только при $s=2$ и $s=1$. В частности, она регулярна при $s=0$. Формально имеем

$$\zeta'_A(0) = -\sum \ln \lambda_n. \qquad (55)$$

Поэтому можно *определить* регуляризованное значение детерминанта оператора A как

$$\det A = \exp(-\zeta'_A(0)). \qquad (56)$$

Дзета-функция может быть связана с ядром $F(x, x', t)$ уравнения теплопроводности или диффузии

$$\frac{\partial F}{\partial t} + A_x F = 0, \qquad (57)$$

где A_x означает, что этот оператор действует на первый аргумент функции F. При начальном условии

$$F(x, x', 0) = \delta(x, x') \qquad (58)$$

функция F описывает диффузию по многообразию M в пятом измерении параметра времени t от точечного источника тепла, находящегося в точке x' в момент $t=0$. Это уравнение теплопроводности изучалось рядом авторов, в том числе Де Виттом [6], Мак-Кином и Синджером [36] и Джилки [25]. Хорошее освещение этого вопроса можно найти в работе [24].

Можно показать, что ядро $F(x, x', t)$ является гладкой функцией x, x' и t при $t>0$, если A — эллиптический оператор. При $t \to 0$ для $F(x, x', t)$ имеется следующее асимптотическое выражение:

$$F(x, x', t) \sim \sum_{n=0}^{\infty} b_n t^{n-2}, \qquad (59)$$

где снова b_n — коэффициенты АМДВ, являющиеся скалярными полиномами по метрике, кривизне и ее ковариантных производных вплоть до порядка $2n$, считая по производным метрики.

Функцию F можно представить через собственные функции и собственные значения оператора A:

$$F(x, x', t) = \sum \phi_n(x)\, \phi_n(x') \exp(-\lambda_n t). \qquad (60)$$

Интегрируя по многообразию, получим

$$Y(t) = \int F(x, x, t)\, (g_0)^{1/2} d^4 x = \sum \exp(-\lambda_n t). \qquad (61)$$

Дзета-функцию можно получить из $Y(t)$ обратным преобразованием Меллина

$$\zeta(s)=\frac{1}{\Gamma(s)}\int_0^\infty Y(t)\,t^{s-1}dt. \tag{62}$$

Используя асимптотическое разложение для F, мы видим, что $\zeta(s)$ имеет полюс при $s=2$ с вычетом B_0 и полюс при $s=1$ с вычетом B_1. Должен бы быть полюс и при $s=0$, но он компенсируется полюсом гамма-функции. Таким образом, $\zeta(0)=B_2$. В определенном смысле полюсы при $s=2$ и $s=1$ соответствуют устранению расходимостей, возникающих из-за первых двух членов в (50).

Если собственные значения известны точно, можно вычислить дзета-функцию и оценить ее производную при $s=0$. В прочих случаях можно получить некоторую информацию из асимптотического разложения ядра уравнения теплопроводности. Предположим, например, что фоновая метрика умножена на постоянный масштабный множитель: $\tilde{g}_0=k^2g_0$; тогда собственные значения λ_n безмассового оператора A будут равны $\tilde{\lambda}_n=k^{-2}\lambda_n$. Поэтому

$$\zeta_{\tilde{A}}(s)=k^{(2s)}\zeta_A(s)$$

и

$$\zeta'_{\tilde{A}}(0)=2\ln k\zeta_A(0)+\zeta'_A(0); \tag{63}$$

следовательно,

$$\ln(\det\tilde{A})=-2\zeta(0)\ln k+\ln(\det A). \tag{64}$$

Поскольку B_2 и, следовательно, $\zeta(0)$, вообще говоря, не равны нулю, мы видим, что интеграл по траекториям не инвариантен относительно конформных преобразований фоновой метрики даже при конформно-инвариантных операторах A. Этот факт известен как конформная аномалия и возникает из-за того, что при регуляризации детерминанта мы вынуждены вводить нормирующую величину μ размерности массы или обратной длины. Альтернативно можно было бы сказать, что мера $D[\phi]=\Pi\mu dy_n$ не является конформно-инвариантной.

Дальнейшие сведения относительно регуляризации через дзета-функцию детерминантов материальных полей можно найти в работах [17, 29, 30, 35].

Регуляризация через дзета-функцию однопетлевого гравитационного члена относительно вакуумного фона была рассмотрена Гиббонсом, Хокингом и Перри [23]. Я коротко изложу эту работу, обобщая ее так, чтобы включить Λ-член.

Член, квадратичный по флуктуациям $\bar{g}$ относительно фоновой метрики g_0, равен

$$I_2[\bar{g}]=\frac{1}{2}\int\bar{g}^{ab}A_{abcd}\bar{g}^{cd}(g_0)^{1/2}d^4x, \tag{65}$$

где

$$g^{ab}=g_0^{ab}+\bar{g}^{ab}, \qquad (66)$$

$$16\pi A_{abcd}=\frac{1}{4}g_{cd}\nabla_a\nabla_b-\frac{1}{4}g_{ac}\nabla_d\nabla_b+\frac{1}{8}(g_{ac}g_{bd}+g_{ab}g_{cd})\nabla_e\nabla^e+$$
$$+\frac{1}{2}R_{ad}g_{bc}-\frac{1}{4}R_{ab}g_{cd}+\frac{1}{16}Rg_{ab}g_{cd}-\frac{1}{8}Rg_{ac}g_{bd}-\frac{1}{8}\Lambda g_{ab}g_{cd}+$$
$$+\frac{1}{4}\Lambda g_{ac}g_{bd}+(a\leftrightarrow b)+(c\leftrightarrow d)+(a\leftrightarrow b,\ c\leftrightarrow d). \qquad (67)$$

Нельзя просто считать однопетлевой член равным $(\det(^1/_2\pi^{-1}\mu^{-1}A))^{1/2}$, поскольку A имеет много нулевых собственных значений, соответствующих тому факту, что действие не меняется при инфинитезимальном диффеоморфизме (калибровочное преобразование)

$$x^a \to x^a+\varepsilon\xi^a,$$
$$g_{ab}\to g_{ab}+2\varepsilon\xi_{(a;\,b)}. \qquad (68)$$

Было бы желательно отфакторизовать калибровочную свободу интегрированием только по калибровочно-неэквивалентным возмущениям $\bar{g}$. Тогда мы получили бы результат, зависящий от детерминанта оператора A на множестве, которое состоит из классов эквивалентности всех полей $\bar{g}$ относительно инфинитезимальных калибровочных преобразований. Как это сделать, было показано Фейнманом [13], Де Виттом [7] и Фаддеевым и Поповым [11]. В действие добавляется член, фиксирующий калибровку:

$$I_f=\frac{1}{2}\int \bar{g}^{ab}B_{abcd}\bar{g}^{cd}(g_0)^{1/2}d^4x. \qquad (69)$$

Оператор B выбирается так, чтобы при любых достаточно малых возмущениях метрики $\bar{g}$, удовлетворяющих требуемым граничным условиям, имелось бы единственное преобразование ξ^a, которое обращается в нуль на границе и удовлетворяет равенству

$$B_{abcd}(\bar{g}^{cd}+2\xi^{(c;\,d)})=0. \qquad (70)$$

Я буду пользоваться гармонической (по фоновой метрике) калибровкой

$$16\pi B_{abcd}=\frac{1}{4}g_{bd}\nabla_a\nabla_c-\frac{1}{8}g_{cd}\nabla_a\nabla_b-\frac{1}{8}g_{ab}\nabla_c\nabla_d+$$
$$+\frac{1}{16}g_{ab}g_{cd}\Box+(a\leftrightarrow b)+(c\leftrightarrow d)+(a\leftrightarrow b,\ c\leftrightarrow d). \qquad (71)$$

Оператор $(A+B)$, вообще говоря, не имеет нулевых собственных значений. Однако $\det(A+B)$ содержит собственные значения произвольно выбранного оператора B. Чтобы скомпенсировать их, необходимо поделить его на детерминант оператора B на подпространстве всех $\bar{g}$, которые являются чистыми калибровками,

т. е. имеют вид $\bar{g}^{ab}=2\xi^{(a;b)}$ при некотором ξ, обращающимся в нуль на границе. Детерминант B на этом подпространстве равен квадрату детерминанта оператора C на пространстве всех векторных полей, которые обращаются в нуль на границе, причем

$$16\pi C_{ab}=-g_{ab}\,\square-R_{ab}. \tag{72}$$

Таким образом, получаем

$$\ln Z=-\hat{I}[g_0]-\frac{1}{2}\ln\det\left(\frac{1}{2}\pi^{-1}\mu^{-2}(A+B)\right)+\ln\det\left(\frac{1}{2}\pi^{-1}\mu^{-2}C\right). \tag{73}$$

Последний член представляет собой так называемый вклад «ду́хов».

Для применения метода дзета-функции необходимо представить $A+B$ как $K-L$, где каждый из операторов K и L имеет только конечное число отрицательных собственных значений. Для этой цели положим

$$A+B=-F+G, \tag{74}$$

где оператор

$$F=-\frac{1}{16}(\nabla_a\nabla^a+2\Lambda) \tag{75}$$

действует на след ϕ метрики $\bar{g}$ (т. е. $\phi=\bar{g}^{ab}g_{0ab}$), а оператор

$$G_{abcd}=-\frac{1}{8}(g_{ac}g_{bd}+g_{ad}g_{bc})\,\nabla^e\nabla_e-\frac{1}{4}(C_{dcab}+C_{dbac})+\frac{1}{6}\Lambda g_{ab}g_{cd} \tag{76}$$

действует на бесследовую часть $\tilde{\phi}$ метрики $\bar{g}$ $(\tilde{\phi}_{ab}=\bar{g}^{ab}-{}^1/_4 g_0^{ab}\phi)$.

Если $\Lambda\leqslant 0$, то оператор F будет иметь только положительные собственные значения. Таким образом, чтобы однопетлевой член был сходящимся, мы должны интегрировать по чисто мнимым ϕ. Это соответствует интегрированию по конформным множителям вида $\Omega=1+i\xi$. При $\Lambda>0$ оператор F будет иметь некоторое конечное число p отрицательных собственных чисел. Поскольку постоянная функция будет собственной функцией F с отрицательным собственным значением (в случае когда нет границы), p по меньшей мере равно единице. Для того чтобы однопетлевой член стал сходиться, мы должны повернуть контур интегрирования коэффициента каждой собственной функции с отрицательным собственным значением так, чтобы этот контур лежал вдоль действительной оси. Это приведет к появлению множителя i^p в Z.

Если фоновая метрика g_0 плоская, то оператор G будет положительно-определенным. Следовательно, интегрирование бесследовых возмущений $\tilde{\phi}$ будет происходить вдоль действительной оси. Это соответствует интегрированию по действительным классам конформной эквивалентности. Однако при неплоских фоновых метриках $\tilde{\phi}$ может иметь некоторое конечное число q отрицательных собственных значений из-за Λ-члена и тензора Вейля. Мы снова

вынуждены повернуть контур интегрирования для этих мод (на этот раз от действительной оси к мнимой); в результате этого в Z появится множитель i^{-q}.

Оператор «дýхов» равен

$$C_{ab} = -g_{ab}(\nabla^e\nabla_e + \Lambda). \quad (77)$$

Если $\Lambda > 0$, то C будет иметь некоторое конечное число r отрицательных собственных значений. Поскольку в Z появляется детерминант C, а не квадратный корень из него, отрицательные собственные значения приведут к множителю $(-1)^r$.

Имеем

$$\ln Z = -\hat{I}[g_0] + \frac{1}{2}\zeta'_F(0) + \frac{1}{2}\zeta'_G(0) - \zeta'_C(0) + \\ + \frac{1}{2}\ln(2\pi\mu^2)(\zeta_F(0) + \zeta_G(0) - 2\zeta_C(0)). \quad (78)$$

Из асимптотического разложения ядра уравнения теплопроводности необходимо получить значения дзета-функций при $s=0$. Из результатов Гиббонса и Перри [21] имеем

$$\zeta_F(0) + \zeta_G(0) - 2\zeta_C(0) = \int \left(\frac{53}{720\pi^2} C_{abcd}C^{abcd} + \frac{763}{540\pi^2}\Lambda^2\right)(g_0)^{1/2}d^4x. \quad (79)$$

Отсюда можно вывести поведение однопетлевого члена при масштабных преобразованиях фоновой метрики. Пусть $\tilde{g}_{0\,ab} = k^2 g_{0\,ab}$; тогда

$$\ln\tilde{Z} = \ln Z + (1-k^2)\hat{I}[g_0] + \frac{1}{2}\gamma\ln k, \quad (80)$$

где γ — правая часть равенства (79). В предположении, что $\hat{I}[g_0] \geqslant 0$, величина $\tilde{Z}$ будет очень мала при больших масштабах, т. е. при больших k. Тот факт, что $\gamma > 0$, означает, что эта величина мала также при очень малых масштабах. Таким образом, в квантовой гравитации может быть обрезание на малых расстояниях. Этот вопрос будет обсуждаться в разд. 10.

7. ФОНОВЫЕ ПОЛЯ

В этом разделе я опишу некоторые положительно-определенные метрики, являющиеся решениями уравнений Эйнштейна, вакуумных или с Λ-членом. В некоторых случаях они представляют собой аналитические продолжения известных лоренцевых решений, хотя их глобальная структура может быть иной. В частности, сечение комплексифицированного многообразия, на котором метрика положительно-определена, может не содержать сингулярностей, имеющихся в лоренцевом сечении. В других случаях эти положительно-определенные метрики могут быть введены только на многообразиях, не имеющих какого-либо сечения, на котором метрика

была бы лоренцевой и действительной. Тем не менее и эти метрики представляют интерес как точки стационарности фазы в определенных интегралах по траектории.

Простейший нетривиальный пример вакуумной метрики — решение Шварцшильда [18, 26]. Обычно оно записывается в виде

$$ds^2 = -\left(1-\frac{2M}{r}\right)dt^2 + \left(1-\frac{2M}{r}\right)^{-1}dr^2 + r^2\,d\Omega^2. \qquad (81)$$

Замена $t=-i\tau$ превращает это выражение в положительно-определенную метрику при $r>2M$. При $r=2M$ имеется явная сингулярность, но она такая же кажущаяся, как сингулярность в начале полярных координат. В этом можно убедиться, введя новую радиальную координату $x=4M(1-2Mr^{-1})^{1/2}$; тогда метрика принимает вид

$$ds^2 = \left(\frac{x}{4M}\right)^2 d\tau^2 + \left(\frac{r^2}{4M^2}\right)^2 dx^2 + r^2 d\Omega^2.$$

Она регулярна при $x=0$, $r=2M$, если τ рассматривается как угловая переменная и по ней производится отождествление с периодом $8\pi M$ (я пользуюсь единицами, в которых гравитационная постоянная $G=1$). Многообразие, задаваемое условиями $x\geqslant 0$, $0\leqslant\tau\leqslant 8\pi M$, называется евклидовым сектором решения Шварцшильда. В нем метрика положительно-определена, асимптотически-плоская и несингулярна (сингулярность кривизны при $r=0$ не лежит в евклидовом секторе).

Ввиду периодичности шварцшильдова решения по мнимому времени с периодом $\beta=8\pi M$ граничная поверхность ∂M радиуса r_0 будет обладать топологией $S^1\times S^2$, и эта метрика будет точкой стационарности фазы в функциональном интеграле для статистической суммы канонического ансамбля с температурой $T=\beta^{-1}=(8\pi M)^{-1}$. Как уже было показано в разд. 2, действие при этом определяется полностью одним только поверхностным членом и равно

$$\hat{I} = \frac{1}{2}\beta M = 4\pi M^2. \qquad (82)$$

Аналогичный евклидов сектор может быть найден для решения Райсснера — Нордстрема с $Q^2+P^2<M^2$, где Q — электрический заряд, а P — магнитный монопольный заряд. В этом случае радиальная координата пробегает значение $r_+\leqslant r<\infty$. Снова внешний горизонт $r=r_+$ будет осью симметрии в плоскости r—τ, а по координате мнимого времени τ производится отождествление с периодом $\beta=2\pi\varkappa^{-1}$, где $\varkappa$ — «гравитационный потенциал» на поверхности внешнего горизонта. Электромагнитное поле F_{ab} в евклидовом секторе будет действительным при мнимом Q и действительном P. В частности, если $Q=iP$, это поле будет самодуальным или антисамодуальным:

$$F_{ab} = \pm\,{}^*F_{ab} = \frac{1}{2}\varepsilon_{abcd}F^{cd}, \qquad (83)$$

где ε_{abcd} — альтернирующий тензор. Если F_{ab} действительно в евклидовом секторе, то операторы, определяющие поведение заряженных полей, будут эллиптическими; поэтому оказывается возможным вычислить однопетлевые члены методом дзета-функции. Затем полученный результат можно аналитически продолжить обратно к действительным Q точно так же, как положительно-определенная метрика аналитически продолжается к лоренцевой.

Поскольку $R=0$, гравитационная часть действия не меняется. Однако имеется еще вклад от электромагнитного лагранжиана $-(1/8\pi)F_{ab}F^{ab}$. Таким образом,

$$\hat{I}=\frac{1}{2}\beta(M-\Phi Q+\psi P), \tag{84}$$

где $\Phi=Q/r_+$ — электростатический потенциал горизонта, а $\psi=P/r_+$ — магнитостатический потенциал.

Подобным же образом можно найти евклидов сектор для метрики Керра при условии, что масса M действительна, а момент количества движения J мнимый. В этом случае метрика будет периодичной в системе отсчета, которая вращается вместе с горизонтом, т. е. точка (τ, r, θ, ϕ) отождествляется с точкой $(\tau+\beta, r, \theta, \phi+i\beta\Omega)$, где Ω — угловая скорость горизонта (Ω мнимая при мнимом J). Как и в случае электромагнитного поля, наилучшим методом представляется вычисление однопетлевых членов с мнимым J и последующее аналитическое продолжение к действительному J. Наличие момента количества движения не влияет на асимптотическую метрику в главном порядке, в котором действие равно

$$\hat{I}=\frac{1}{2}\beta M, \quad \beta=2\pi\varkappa^{-1},$$

где $\varkappa$ — гравитационный потенциал поверхности.

Метрики Тауба — НУТ [32, 37] составляют другой интересный класс вакуумных решений. Их можно рассматривать как гравитационные «дионы» с обычной массой M «электрического» типа и гравитационной массой N «магнитного» типа. Соответствующая метрика может быть представлена в виде

$$ds^2=-V\left(dt+4N\sin^2\frac{\theta}{2}d\phi\right)^2+V^{-1}dr^2+(r^2+N^2)(d\theta^2+\sin^2\theta\, d\phi^2), \tag{85}$$

где $V=1-(2Mr+N^2)/(r^2+N^2)$. Эта метрика регулярна на полуоси $\theta=0$, но имеет сингулярность при $\theta=\pi$, поскольку член $\sin^2(\theta/2)$ в метрике означает, что малый контур вокруг этой оси не стягивается в точку при $\theta=\pi$. Эту сингулярность можно рассматривать как аналог дираковской струны в электродинамике, возникающей при наличии монопольного магнитного заряда. Такую сингулярность можно устранить путем введения новой координаты

$$t'=t+4N\phi. \tag{86}$$

Тогда метрика принимает вид

$$ds^2 = -V\left(dt' - 4N\cos^2\frac{\theta}{2}d\phi\right)^2 + V^{-1}dr^2 + (r^2+N^2)(d\theta^2 + \sin^2\theta d\phi^2). \tag{87}$$

Она регулярна при $\theta=\pi$, но сингулярна при $\theta=0$. Следовательно, можно использовать координаты (t, r, θ, ϕ) для покрытия окрестности северного полюса ($\theta=0$) и координаты (t', r, θ, ϕ) для покрытия окрестности южного полюса ($\theta=\pi$). Поскольку по ϕ происходит отождествление с периодом 2π, из формулы (86) следует, что по t и t' происходит отождествление с периодом $8\pi N$. Поэтому если ψ — регулярное поле с зависимостью от t вида $\exp(-i\omega t)$, то ω должна удовлетворять условию

$$4N\omega \text{ равно целому числу.} \tag{88}$$

Оно является аналогом дираковского условия квантования и связывает «магнитный» заряд N решения Тауба — НУТ с «электрическим» зарядом или энергией ω поля ψ. Процесс устранения сингулярности дираковской струны через введение координат t и t' и периодическое отождествление превращает топологию поверхностей постоянного r из $S^2\times R^1$ в S^3; на них ($t/2N$), θ и ϕ — угловые координаты Эйлера.

Метрика (85) также имеет сингулярности при $V=0$ и $V=\infty$. Как и в случае Шварцшильда, $V=\infty$ соответствует неустранимой сингулярности кривизны, но $V=0$ соответствует некоторому горизонту, и эта сингулярность может быть устранена периодическим отождествлением мнимой координаты времени. Это отождествление совместимо с тем отождествлением, которое устраняет дираковскую струну, если оба периода равны, что имеет место при $N=\pm iM$. Если это так, то при действительном M метрика действительна и положительно определена в области $r>M$, а кривизна самодуальна или антисамодуальна:

$$R_{abcd} = \pm{}^*R_{abcd} = \pm\frac{1}{2}\varepsilon_{abef}R^{ef}{}_{cd}. \tag{89}$$

Кажущаяся сингулярность при $r=M$ становится единственной точкой — началом гиперсферических координат, в чем можно убедиться, вводя новые радиальные и временные переменные

$$x = 2(2M(r-M))^{1/2}, \qquad \psi = -\frac{it}{2M}. \tag{90}$$

Метрика при этом принимает вид

$$ds^2 = \frac{Mx^2}{2(r+M)}(d\psi + \cos\theta\, d\phi)^2 + \frac{r+M}{2M}dx^2 + \frac{x^2(r+M)}{8M}(d\theta^2 + \sin^2\theta d\phi^2). \tag{91}$$

Таким образом, многообразие, которое определяется условиями $x \geqslant 0$, $0 \leqslant \psi \leqslant 4\pi$, $0 \leqslant \theta \leqslant \pi$, $0 \leqslant \phi \leqslant 2\pi$, причем ψ, θ, ϕ интерпретируются как гиперсферические углы Эйлера, топологически представляет собой R^4 и обладает несингулярной положительно-определенной метрикой. Эта метрика — асимптотически-плоская в том смысле, что тензор Римана убывает как r^{-3} при $r \to \infty$, но не является асимптотически-евклидовой, поскольку для этого кривизна должна быть пропорциональной r^{-4}. Поверхности постоянного r топологически суть S^3, но их метрика — это метрика деформированной сферы. Орбиты вектора Киллинга $\partial/\partial\psi$ задают расслоение Хопфа π: $S^3 \to S^2$, где S^2 параметризуется координатами θ и ϕ. Индуцированная на S^2 метрика будет метрикой двумерной сферы радиуса $(r^2 - M^2)^{1/2}$, тогда как слои будут окружностями длины $8\pi M V^{1/2}$. Таким образом, граница при большом радиусе есть в некотором смысле $S^1 \times S^2$, но является скрученным (twisted) произведением.

Имеется также возможность скомбинировать самодуальные решения Тауба — НУТ [29, 30]. Причина состоит в том, что притяжение между массами электрического типа M уравновешивается отталкиванием между мнимыми массами магнитного типа N. Соответствующая метрика имеет вид

$$ds^2 = U^{-1}(d\tau + \omega \cdot d\mathbf{x})^2 + U\, d\mathbf{x} \cdot d\mathbf{x}, \tag{92}$$

где

$$U = 1 + \sum \frac{2M_i}{r_i}, \quad \operatorname{rot} \omega = \operatorname{grad} U. \tag{93}$$

Здесь r_i — расстояние от i-го решения НУТ в плоской трехмерной метрике $d\mathbf{x} \cdot d\mathbf{x}$. Операции rot и grad, как и вектор v, относятся к этой трехмерной метрике. Для каждого решения НУТ $N_i = iM_i$.

Векторные поля ω будут иметь сингулярности дираковской струны от каждого решения НУТ. Если массы M_i все равны между собой, эти струнные сингулярности и сингулярности типа горизонтов при $r_i = 0$ все могут быть устранены отождествлением с периодом $8\pi M$. Граничная поверхность при больших радиусах будет при этом поверхностью, подобной линзе [41]. Топологически это поверхность S^3 с отождествлением n точек в слое S^1 расслоения Хопфа $S^3 \to S^2$, где n — число решений НУТ.

Эту граничную поверхность даже локально нельзя вложить в плоское пространство, поэтому мы не можем получить поправочный член K^0 для действия. Если пытаться «почти» вложить ее настолько, насколько это возможно, то для действия получается значение $4\pi n M^2$, т. е. такое же, как в случае Шварцшильда с $n = 1$ [4]. Фактически присутствие гравитационной магнитной массы изменяет топологию пространства и не позволяет ей быть асимптотически-плоской в обычном смысле. Но можно получить асимптотически-плоское пространство, содержащее одинаковое число

n решений НУТ ($N=iM$) и антиНУТ ($N=-iM$). Поскольку решения НУТ и антиНУТ взаимно притягиваются, они должны удерживаться на расстоянии друг от друга электромагнитным полем. Такое решение будет фактически решением Израэля — Вильсона [27, 33]. Гравитационная часть действия равна при этом $8\pi n M^2$, так что каждое НУТ и антиНУТ дает вклад $4\pi M^2$.

Теперь я перейду к положительно-определенным метрикам, которые являются решениями уравнений Эйнштейна с Λ-членом на компактных многообразиях без края.

Простейшим примером служит S^4 с метрикой, индуцированной вложением этого многообразия как сферы радиуса $(3\Lambda^{-1})^{1/2}$ в пятимерное евклидово пространство. При этом мы имеем аналитическое продолжение пространства де Ситтера [19]. Метрика может быть выражена через вектор Киллинга $\partial/\partial\tau$:

$$ds^2=\left(1-\frac{1}{3}\Lambda r^2\right)d\tau^2+\left(1-\frac{1}{3}\Lambda r^2\right)^{-1}dr^2+r^2\,d\Omega^2. \qquad (94)$$

Она содержит сингулярность типа горизонта при $r=(3\Lambda^{-1})^{1/2}$. Фактически это двумерная сфера площади $12\pi\Lambda^{-1}$, которая является геометрическим местом нулей вектора Киллинга $\partial/\partial\tau$. Действие равно $-3\pi\Lambda^{-1}$.

Можно также получить решения с черной дырой, которые асимптотически не плоски, а деситтеровы. Простейшим из них является решение Шварцшильда — де Ситтера [19]. Соответствующая метрика имеет вид

$$ds^2=V d\tau^2+V^{-1}dr^2+r^2\,d\Omega^2, \qquad (95)$$

где

$$V=1-2Mr^{-1}-\frac{1}{3}\Lambda r^2.$$

Если $\Lambda<(9M^2)^{-1}$, то имеются два положительных значения r, при которых $V=0$. Меньшее из них соответствует горизонту черной дыры, тогда как большее аналогично «космологическому горизонту» в пространстве де Ситтера. Кажущуюся сингулярность на каждом из горизонтов можно устранить периодическим отождествлением по τ. Но периоды, требующиеся для этих двух горизонтов, различны, за исключением предельного случая $\Lambda=(9M^2)^{-1}$. В этом случае многообразие представляет собой $S^2\times S^2$ с метрикой прямого произведения, а действие равно $-2\pi\Lambda^{-1}$.

Можно получить также решение Керра — де Ситтера [19]. При мнимом моменте количества движения метрика будет положительно определена для значений r, лежащих между космологическим горизонтом и внешним горизонтом черной дыры. Сингулярности типа горизонта снова можно устранить периодическим отождествлением, и при специальном выборе параметров периоды будут соизмеримы [38]. В этом случае мы получим метрику без

сингулярностей на S^2-расслоении над S^2. При этом действие равно —0,9553 $(2\pi\Lambda^{-1})$.

Можно получить и решения Тауба — де Ситтера. В дополнение к обычным сингулярностям Тауба — НУТ они будут обладать еще космологическим горизонтом. Все горизонты и сингулярности дираковской струны могут быть устранены одновременно в том предельном случае, когда многообразие представляет собой CP^2 (комплексное двумерное проективное пространство с обычной келеровой метрикой) [20]. Действие равно $-{}^9/_4\pi\Lambda^{-1}$.

Можно получить решения, являющиеся произведением двух двумерных пространств постоянной кривизны [17]. О случае $S^2\times S^2$ уже упоминалось, но есть также тривиальный случай плоского тора $T^2\times T^2$. В других примерах оба пространства имеют жанры $g_1>1$ и $g_2>1$ и Λ-член должен быть отрицательным. Действие равно $-(2\pi/\Lambda)(g_1-1)(g_2-1)$.

Наконец, для завершения этого каталога положительно-определенных решений уравнений Эйнштейна нужно упомянуть $K3$. Это компактное четырехмерное многообразие, которое может быть реализовано как поверхность четвертого порядка в трехмерном комплексном проективном пространстве CP^3. Это многообразие может быть снабжено положительно-определенной метрикой, кривизна которой самодуальна, вследствие чего эта метрика будет решением уравнений Эйнштейна с $\Lambda=0$ [46]. Более того, с точностью до отождествлений $K3$ является единственным компактным многообразием, допускающим самодуальную метрику. Действие равно нулю.

Компактные четырехмерные многообразия обладают топологическими инвариантами, которые могут быть представлены в виде интегралов от кривизны:

$$\chi=\frac{1}{128\pi^2}\int R_{abcd}R_{efgh}\varepsilon^{abef}\varepsilon^{cdgh}\,(g)^{1/2}d^4x, \qquad (96)$$

$$\tau=\frac{1}{96\pi^2}\int R_{abcd}R^{ab}{}_{ef}\varepsilon^{cdef}\,(g)^{1/2}d^4x; \qquad (97)$$

χ есть число Эйлера данного многообразия и равно альтернированной сумме чисел Бетти:

$$\chi=B_0-B_1+B_2-B_3+B_4. \qquad (98)$$

Число Бетти B_p есть число независимых замкнутых p-мерных поверхностей, которые не являются границей какой-либо $p+1$-мерной поверхности. Оно равно также числу независимых гармонических p-форм. Для замкнутых многообразий $B_n=B_{4-p}$ и $B_0=B_4=1$. Если многообразие односвязно, то $B_1=B_3=0$ и, следовательно, $\chi\geqslant 2$.

Сигнатура Хирцбруха τ интерпретируется следующим образом: B_2 гармонических 2-форм можно разделить на B_2^+ самодуальных и B_2^- антисамодуальных 2-форм. Тогда $\tau=B_2^+-B_2^-$ опреде-

ляет гравитационный вклад в аномалию аксиального тока [10, 29, 30, 32].

Для S^4 имеем $\chi=2$ и $\tau=0$; для CP^2 $\chi=3$, $\tau=1$; для S^2-расслоения над S^2 $\chi=4$, $\tau=0$; для $K3$ $\chi=24$, $\tau=16$; для произведения двумерных пространств с жанрами g_1, g_2 имеем $\chi=4(g_1-1)(g_2-1)$, $\tau=0$.

В некомпактном случае в формулах для χ и τ имеются добавочные поверхностные члены. Евклидово пространство и самодуальное решение Тауба — НУТ имеют $\chi=1$, $\tau=0$, а для решения Шварцшильда $\chi=2$, $\tau=0$.

8. ГРАВИТАЦИОННАЯ ТЕРМОДИНАМИКА

Как уже было объяснено в разд. 3, статистическая сумма

$$Z=\sum \exp(-\beta E_n)$$

для системы с температурой $T=\beta^{-1}$, находящейся в сферической полости радиуса r_0, задается интегралом по траекториям по всем метрикам, которые могут быть введены внутри границы ∂M с топологией $S^2\times S^1$, где S^2— сфера радиуса r_0, а S^1 — окружность длиной β. В приближении стационарной фазы, описанном в разд. 5, преобладающий вклад вносят метрики, близкие к классическому решению с данными граничными условиями. Одним из таких решений будет плоское пространство, в котором по евклидовой координате времени производится отождествление с периодом β. Топологически это $R^3\times S^1$. Действие фоновой метрики равно нулю, таким образом, она не дает вклада в логарифм статистической суммы. Если пренебречь малыми поправками, возникающими из-за конечных размеров полости, то можно точно вычислить однопетлевой член Z_g:

$$\ln Z_g=\frac{4\pi^5 r_0^3 T^3}{135}. \tag{99}$$

Его можно интерпретировать как статистическую сумму тепловых гравитонов на фоне плоского пространства.

Метрика Шварцшильда при $M=(8\pi T)^{-1}$ представляет собой другое решение, удовлетворяющее тем же граничным условиям. Это решение обладает топологией $R^2\times S^2$, и его действие $\hat{I}=\beta^2/16\pi=4\pi M^2$. Однопетлевой член пока еще не вычислен, но по масштабным соображениям, приведенным в разд. 6, он должен иметь вид

$$\frac{106}{45}\ln\left(\frac{\beta}{\beta_0}\right)+f(r_0\beta^{-1}), \tag{100}$$

где постоянная β_0 связана с нормировочной постоянной μ. Если $r_0\beta^{-1}\gg 1$, то размеры полости много больше черной дыры и следует ожидать, что значение $f(r_0\beta^{-1})$ приближается к значению для плос-

кого пространства (99). Поэтому функция f должна иметь вид

$$f(r_0\beta^{-1})=\frac{4\pi^5 r_0^3}{135\beta^3}+O(r_0^2\beta^{-2}). \quad (101)$$

По этой статистической сумме можно вычислить среднее значение энергии

$$\langle E\rangle=\frac{\sum E_n \exp(-\beta E_n)}{\exp(-\beta E_n)}=-\frac{\partial}{\partial\beta}\ln Z. \quad (102)$$

Применяя это выражение к вычислению вклада $(-\beta^2/16\pi)$ в $\ln Z$ от действия решения Шварцшильда, мы получим $\langle E\rangle=M$, как и следовало ожидать. Можно вычислить также энтропию, которая может быть определена как

$$S=-\sum p_n \ln p_n, \quad (103)$$

где $p_n=Z^{-1}\exp(-\beta E_n)$ — вероятность того, что система находится в n-м состоянии. Тогда

$$S=\beta\langle E\rangle+\ln Z. \quad (104)$$

Применяя эту формулу к вычислению вклада от действия для метрики Шварцшильда, получим

$$S=4\pi M^2=\frac{1}{4}A, \quad (105)$$

где A — площадь горизонта событий.

Это замечательный результат, так как он показывает, что в дополнение к энтропии, возникающей от однопетлевого члена (ее можно рассматривать как энтропию тепловых гравитонов на шварцшильдовом фоне), черные дыры обладают внутренней энтропией, возникающей от действия той метрики, которая соответствует стационарности фазы. Эта внутренняя энтропия в точности совпадает с энтропией, приписываемой черной дыре на основании вычислений рождения частиц на заданном фоне и использования первого закона механики черных дыр (см. [48, 49]). Отсюда мы видим, что идея о привнесении гравитацией в физику нового уровня непредсказуемости или хаотичности подтверждается не только полуклассическим приближением, но и подходом, в котором гравитационное поле квантовано.

Одной из причин, по которым классические решения в теории гравитации обладают внутренней энтропией, а в теории Янга — Миллса или $\lambda\phi^4$ не обладают, является то, что действия этих теорий в отличие от гравитационного действия масштабно-инвариантны. Если g_0— асимптотически-плоское решение с периодом $k\beta$ и действием $\hat{I}[g_0]$, то $k^2 g_0$ — решение с периодом $k\beta$ и действием $k^2\hat{I}$. Это значит, что действие $\hat{I}$ должно иметь вид $c\beta^2$, где c — постоянная, которая зависит от топологии этого решения. Тогда $\langle E\rangle=2c\beta$, $\beta\langle E\rangle=$ $=2c\beta^2$ и в то же время $\ln Z=-\hat{I}=-c\beta^2$. Отсюда $S=c\beta^2$. Причиной

того, что действие $\hat{I}$ равно $^1/_2\beta\langle E\rangle$, а не $\beta\langle E\rangle$, как следовало бы ожидать для единственного состояния с энергией $\langle E\rangle$, является отличие топологии решения Шварцшильда от топологии периодически отождествленного плоского пространства. Из того факта, что число Эйлера для решения Шварцшильда равно 2, следует, что вектор Киллинга трансляции по времени $\partial/\partial\tau$ должен быть равен нулю на некотором множестве (в действительности на двумерной сфере). Таким образом, поверхности постоянного τ имеют две границы: одну на поверхности сферической полости радиуса r_0 и другую на горизонте $r=2M$.

Рассмотрим теперь область шварцшильдова решения, ограниченную поверхностями $\tau=\tau_1$, $\tau=\tau_2$ и $r=r_0$ (рис. 4). Амплитуда $\langle\tau_2|\tau_1\rangle$ перехода от поверхности τ_1 к поверхности τ_2 определяется интегралом по траекториям по всем метрикам, которые могут быть введены в пределах этой границы, причем главный вклад вносит метрика,

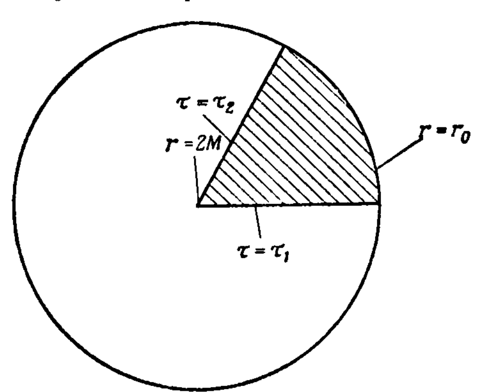


Рис. 4. Плоскость (τ, r) решения Шварцшильда.
Амплитуда $\langle\tau_2|\tau_1\rangle$ перехода с поверхности τ_1 к поверхности $\tau=\tau_2$ определяется преимущественно действием заштрихованной части решения Шварцшильда.

обеспечивающая стационарность фазы, а ею как раз и является часть решения Шварцшильда, ограниченная этими поверхностями. Действие для этой метрики стационарности фазы дают поверхностные члены, поскольку $R=0$. Поверхностные члены для поверхностей $\tau=\tau_1$ и $\tau=\tau_2$ взаимно сокращаются. Вклад поверхности $r=r_0$ равен $^1/_2M(\tau_2-\tau_1)$, но, кроме того, будет также вклад «угла» при $r=2M$, где встречаются поверхности $\tau=\tau_1$ и $\tau=\tau_2$, так как вторая фундаментальная форма границы K ведет себя здесь как δ-функция. Слегка закругляя этот угол, мы можем вычислить этот вклад; он оказывается равным $^1/_2M(\tau_2-\tau_1)$. Таким образом, полное действие равно $\langle E\rangle(\tau_2-\tau_1)$ и $\langle\tau_2|\tau_1\rangle=\exp(-\langle E\rangle(\tau_2-\tau_1))$, как и следовало ожидать для единственного состояния с энергией $E=\langle E\rangle$. Но если рассматривать статистическую сумму, то границей будет просто $r=r_0$ и, следовательно, действие будет равно $^1/_2\beta E$, а не βE. Эта разность, равная $^1/_4A$, дает энтропию черной дыры.

Таким образом, мы видим: тот факт, что гравитационное поле может иметь различные топологии, приводит к качественно новым явлениям. Эти явления нельзя было бы обнаружить, используя канонический подход, поскольку такие метрики, как решение Шварцшильда, в этом подходе не допускаются.

Приведенный вывод статистической суммы и энтропии черной дыры основан на использовании канонического ансамбля, в котором система находится в равновесии с бесконечным резервуаром энергии при температуре T. Однако канонический ансамбль неустойчив в присутствии черной дыры, поскольку если черная дыра поглотит хоть немного дополнительной энергии, она несколько остынет и будет и далее поглощать энергии больше, чем испускать ее. Эта патология отражается в том факте, что $\langle \Delta E^2 \rangle = \langle E^2 \rangle - \langle E \rangle^2 = (1/Z)(\partial^2 Z/\partial\beta^2) = (\partial \ln Z/\partial\beta)^2 = -1/8\pi$, т. е. отрицательна. Для получения разумных результатов в отношении черных дыр мы должны пользоваться микроканоническим ансамблем, в котором в изолированную полость заключается определенное количество энергии E и рассматриваются все возможные конфигурации, которые обладают заданной энергией. Пусть $N(E)dE$ — число состояний гравитационного поля с энергиями между E и $E+dE$ в сферической полости r_0. Статистическую сумму дает преобразование Лапласа над $N(E)$:

$$Z(\beta) = \int_0^\infty N(E) \exp(-\beta E)\, dE. \tag{106}$$

Отсюда плотность состояний формально получается обратным преобразованием Лапласа:

$$N(E) = \frac{1}{2\pi i} \int_{-i\infty}^{i\infty} Z(\beta) \exp(E\beta)\, d\beta. \tag{107}$$

Для больших β главный вклад в $Z(\beta)$ дает действие метрики Шварцшильда, и этот вклад имеет вид $\exp(-\beta^2/16\pi)$. Следовательно, правая часть (107) будет расходиться, если интеграл взят по мнимой оси β, как это предполагается. Чтобы получить конечное значение для (107), приходится придерживаться иного способа: брать интеграл вдоль действительной оси β. Это очень похоже на процедуру, использованную при вычислении функционального интеграла в приближении стационарной фазы, где мы поворачивали контур интегрирования для каждого квадратичного члена так, чтобы получался сходящийся гауссов интеграл. При следовании этому рецепту множитель $1/2\pi i$ в (107) дает мнимое значение для плотности состояний $N(E)$, если статистическая сумма $Z(\beta)$ действительна. Однако, как уже упоминалось в разд. 6, оператор G, которому подчиняются неконформные или бесследовые возмущения в метрике Шварцшильда, обладает одним отрицательным собственным значением. Это вносит в однопетлевой член для Z множитель i. Таким образом, статистическая сумма оказывается чисто мнимой, а плотность сос-

тояний — действительной. Это как раз то, чего и следовало ожидать. Из-за того что канонический ансамбль не определен корректно, получаем патологическую статистическую сумму, но вследствие правильного поведения микроканонического ансамбля плотность состояний действительна и положительна.

Выходить за пределы приближения стационарной фазы при вычислении интеграла в (107) не имеет смысла, поскольку статистическая сумма Z вычислена только в этом приближении. Если мы возьмем в качестве вклада фоновой метрики $\exp(-\beta^2/16\pi)$, то получим, что черная дыра массы M имеет плотность состояний $N(M) = 2\pi^{-1/2}\exp(4\pi M^2)$. Следовательно, интеграл в (106) не сходится, если не повернуть контур интегрирования так, чтобы он лежал вдоль мнимой оси E. Если включить однопетлевой член Z_g, то точка стационарности фазы при интегрировании по β в (107) для плоской фоновой метрики находится при

$$E = \frac{-\partial \ln Z_g}{\partial\beta} \tag{108}$$

и для фоновой метрики Шварцшильда при

$$E = \frac{\beta}{8\pi} - \frac{\partial \ln Z_g}{\partial\beta}. \tag{109}$$

Этим уравнениям можно дать такую интерпретацию: E равно энергии тепловых гравитонов и черной дыры, если она есть. Используя приближенный вид Z_g, находим, что при объеме полости V, удовлетворяющем неравенству

$$E^5 < \frac{\pi^2}{15}(8354{,}5)V, \tag{110}$$

преобладающий вклад в N дает фоновая метрика плоского пространства. Таким образом, в этом случае наиболее вероятное состояние системы — тепловые гравитоны без каких-либо черных дыр. Если V меньше, чем нужно для выполнения условия (110), то появляются две точки стационарности фазы для фоновой метрики Шварцшильда. Одна из них с меньшим значением β дает вклад, больший чем фоновая плоская метрика. Следовательно, наиболее вероятным состоянием системы будет черная дыра в равновесии с тепловыми гравитонами. Эти результаты подтверждают прежние выводы, основанные на полуклассических приближениях [20, 28].

9. ЗА ПРЕДЕЛАМИ ОДНОПЕТЛЕВОГО ПРИБЛИЖЕНИЯ

В разд. 5 действие было разложено в ряд Тейлора в окрестности фонового поля, которое было решением классических полевых уравнений. Интеграл по траекториям был взят по квадратичным членам, а членами более высокого порядка мы пренебрегли. В перенормируемых теориях, таких, как квантовая электродинамика, тео-

рия Янга — Миллса, $\lambda\phi^4$, с помощью дифференциального оператора A, присутствующего в квадратичной («свободной») части действия, можно вычислить и эти члены более высокого порядка (члены «взаимодействия»). Их вклад можно представить диаграммами Фейнмана с двумя или более замкнутыми петлями, где линии диаграмм соответствуют пропагатору или функции Грина A^{-1}, а вершины — членам взаимодействия, скажем кубичным, если в вершине сходятся три линии, и т. д. В этих перенормируемых теориях неопределенные величины, возникающие при регуляризации петель второго и более высокого порядков, оказываются связанными с неопределенной нормировочной величиной μ для одной петли. Поэтому все эти величины могут быть включены в переопределение постоянной взаимодействия и каких-либо масс, присутствующих в теории.

В квантовой гравитации ситуация совсем иная. Однопетлевой член относительно плоской или топологически тривиальной вакуумной метрики не содержит нормировочной постоянной μ. Но в окрестности топологически нетривиального фона мы получаем $\ln Z_g$, пропорциональный $(106/45)\chi \ln \mu$, где Z_g — однопетлевой член, а χ — число Эйлера. Это можно представить как добавление к действию эффективного топологического члена $-k(\mu)\chi$, где $k(\mu)$ — масштабно-неинвариантная топологическая постоянная связи. Вообще говоря, мы не можем обеспечить такую же топологическую интерпретацию для зависимости от μ однопетлевого члена относительно фоновой метрики, которая является решением полевых уравнений с ненулевыми материальными полями. Однако это может быть сделано в специальном случае, когда материальные поля связаны с гравитационным полем локальной суперсимметрией или зависящими от спиноров калибровочными преобразованиями. Таковы различного рода теории супергравитации и расширенной супергравитации [4, 15].

Двухпетлевые члены в супергравитации и, возможно, также в чистой гравитации, по-видимому, не вносят каких-либо дополнительных неопределенных величин. Но кажется весьма вероятным, что и в супергравитации, и в чистой гравитации в трехпетлевых и более высоких приближениях возникнут добавочные неопределенности, хотя вычисления, необходимые для выяснения этого, столь громоздки, что никто не пытался их проделать. Даже если каким-то чудом никаких добавочных неопределенностей в многопетлевых членах не возникнет, у нас все же не будет разумной процедуры для вычисления интеграла по траекториям, так как разложение по возмущениям в окрестности заданного фонового поля в теории гравитации имеет весьма узкие рамки применимости в отличие от перенормируемых теорий вроде теории Янга — Миллса или $\lambda\phi^4$. В последней теории квадратичный («свободный») член в действии $\int (\Delta\phi)^2 d^4x$ ограничивает член «взаимодействия» $\lambda \int \phi^4 d^4x$. Это озна-

чает, что среднее значение члена взаимодействия может быть вычислено по мере $D[\phi]\exp(-\int(\Delta\phi)^2 d^4x)$ или, иными словами, с помощью диаграмм Фейнмана, в которых линии соответствуют свободному пропагатору. Аналогично и в квантовой электродинамике или теории Янга — Миллса член взаимодействия только третьего или четвертого порядка по полям и ограничен свободным членом. В гравитационном же случае разложение Тейлора в окрестности фоновой метрики содержит члены взаимодействия всех порядков по возмущениям метрики и квадратичные члены по производным этих возмущений. Эти члены взаимодействия не ограничены свободным квадратичным членом, поэтому их средние значения по мере, заданной этим квадратичным членом, не определены. Иными словами, изображать их диаграммами Фейнмана высших порядков не имеет никакого смысла. Это не будет сюрпризом для тех, кто работает в классической теории относительности, а не в квантовой теории поля. Мы знаем, что объект, подобный черной дыре, нельзя представить в виде возмущения плоского пространства.

В классической общей теории относительности проблему ограниченной применимости теории возмущений можно пытаться решать путем подбора асимптотических разложений относительно различных фоновых метрик. Поэтому представляется естественным попытаться проделать нечто подобное в квантовой гравитации. Для того чтобы гарантировать калибровочную инвариантность, по-видимому, необходимо, чтобы эти фоновые метрики были решениями классических уравнений поля. Насколько нам известно, при заданной топологии и заданных граничных условиях имеется только одно решение полевых уравнений или в крайнем случае такие решения составляют конечномерное семейство. Следовательно, решения заданной топологии не могут быть плотны в пространстве метрик этой топологии. Однако эйнштейновское действие в отличие от янг-миллсовского, по-видимому, не приводит к какому-либо барьеру в переходах между полями разной топологии.

Одним из путей убедиться в этом является использование исчисления Редже [40]. При этом подходе пространственно-временнóе многообразие разлагается на симплициальный комплекс. Каждый 4-симплекс берется плоским и определяется длинами его ребер (т. е. 1-симплексов). Однако углы между гранями (т. е. 2-симплексами), вообще говоря, таковы, что 4-симплексы нельзя собрать в плоское четырехмерное пространство. Таким образом, имеется дисторсия, которая может быть представлена как δ-функции по кривизне, сконцентрированные на гранях. Полное действие равно $(-1/8\pi)\Sigma A_i\delta_i$; сумма берется по всем 2-симплексам, причем A_i — площадь i-го симплекса, а δ_i — угловой эффект при этом симплексе, т. е. δ_i равен 2π минус сумма углов между теми 3-симплексами, которые связаны данным 2-симплексом.

Комплекс, в котором действие стационарно при малых вариациях длин ребер, можно рассматривать как дискретную аппроксимацию к гладкому решению уравнений Эйнштейна. Однако на исчисление Редже можно также смотреть как на определение действия для определенного класса метрик без какого-либо приближения. Это действие останется корректно определенным и конечным, даже если длины ребер будут выбраны так, что некоторые из симплексов стянутся в симплексы меньших размерностей. Например, если a, b, c — длины сторон треугольника (2-симплекса), то они должны удовлетворять неравенствам $a<b+c$ и т. д. Если $a=b+c$, этот 2-симплекс стягивается в 1-симплекс. В общем случае симплициальный комплекс перестает быть многообразием, если некоторые из его симплексов стягиваются до меньших размерностей. Однако действие по-прежнему хорошо определено. При этом мы можем «раздуть» некоторые из симплексов так, чтобы получилось новое многообразие с иной топологией. Таким способом можно непрерывно переходить от одной метрической топологии к другой.

Отсюда возникает идея, что могут быть квантовые флуктуации метрики не только в пределах отдельно взятой топологии, но и от одной топологии к другой. На эту возможность впервые указал Уилер [45], который выдвинул гипотезу, что в масштабах планковской длины пространство-время может иметь «пеноподобную» структуру. В следующем разделе я попытаюсь изложить математический аппарат для описания этой пеноподобной структуры. Существует надежда, что при рассмотрении метрик всех возможных топологий будет обнаружено, что классические решения в некотором смысле плотны в пространстве всех метрик. Тогда можно надеяться представить интегралы по траекториям как суммы фоновых и однопетлевых членов для этих решений. Можно было бы также рассчитывать на выделение некоторого конечного числа решений, которые дают главный вклад.

10. ПРОСТРАНСТВЕННО-ВРЕМЕННАЯ ПЕНА

Итак, мы хотели бы знать, какие топологии метрик, обеспечивающих стационарность фазы, дают главный вклад в функциональный интеграл. Для этого удобно рассмотреть интеграл по траекториям по всем компактным метрикам, при которых пространственно-временно́й объем имеет данное значение V. Это не означает, что мы в самом деле считаем пространство-время компактным. Мы лишь пользуемся общепринятой схемой нормировки типа периодических граничных условий в обычной квантовой теории поля: сначала мы работаем с конечным объемом, чтобы иметь конечное число состояний, а затем рассматриваем значения различных величин в пересчете на единичный объем в пределе, когда объем устремляется к бесконечности.

Чтобы рассматривать интеграл по траекториям по метрикам с данным четырехмерным объемом V, в действие вводится член $\Lambda V/8\pi$,

где Λ нужно понимать как множитель Лагранжа (коэффициент $1/8\pi$ взят для удобства). Этот член имеет такой же вид, как космологический член, но мотивировка его введения и его численное значение совершенно иные: из наблюдательных данных следует, что любое космологическое Λ должно иметь столь малое значение, что практически им можно пренебречь, тогда как значение множителя Лагранжа оказывается очень большим — порядка 1 в планковских единицах.

Пусть

$$Z[\Lambda]=\int D[g]\exp\left(-\hat{I}[g]-\frac{\Lambda}{8\pi}V[g]\right), \qquad (111)$$

где интеграл берется по всем метрикам на некотором компактном многообразии. $Z[\Lambda]$ можно интерпретировать как «статистическую сумму» для того, что я буду называть *объемным каноническим ансамблем*, т. е.

$$Z[\Lambda]=\sum_n\left\langle\phi_n\right|\exp\left(-\frac{\Lambda V}{8\pi}\right)\left|\phi_n\right\rangle, \qquad (112)$$

где сумма берется по всем состояниям $|\phi_n\rangle$ гравитационного поля. Зная $Z[\Lambda]$, можно вычислить $N(V)dV$ — число гравитационных полей с 4-объемами между V и $V+dV$:

$$N(V)=\frac{1}{16\pi^2 i}\int_{-i\infty}^{i\infty}Z[\Lambda]\exp(\Lambda V)\,d\Lambda. \qquad (113)$$

В выражении (113) контур интегрирования должен быть взят правее любых сингулярностей $Z[\Lambda]$ на мнимой оси.

Нам нужно сравнить вклады в N от различных топологий. Подходящей мерой сложности топологии является число Эйлера χ. В случае односвязных многообразий χ и сигнатура τ, по-видимому, характеризуют многообразие с точностью до гомотопий и, возможно, до гомеоморфизмов, хотя это не доказано. В случае многосвязных многообразий возможность классификации отсутствует, так как нет никакого алгоритма для ответа на вопрос, являются ли два многосвязных четырехмерных многообразия гомеоморфными или гомотопными. Такое положение представляется серьезным основанием для того, чтобы ограничиться рассмотрением только односвязных многообразий. Другим основанием может служить то, что многосвязное многообразие всегда можно развернуть. Это может привести к некомпактному многообразию, но следует ожидать, что его можно будет заключить в большой объем V так, чтобы действие на единицу объема изменилось лишь незначительно.

В приближении стационарной фазы можно ожидать, что главный вклад в интеграл по траекториям Z дадут метрики, близкие к решениям уравнений Эйнштейна с Λ-членом. Из масштабного поведе-

ния действия следует, что для таких решений

$$\Lambda = -8\pi c V^{-1/2}, \quad \hat{I} = -\frac{8\pi c^2}{\Lambda}, \tag{114}$$

где c — постоянная (положительная или отрицательная), которая зависит от выбора решения и топологии, а действие $\hat{I}$ включает теперь Λ-член. Постоянная c имеет нижнюю границу $-(^3/_8)^{1/2}$, которая соответствует ее значению для S^4. Верхний предел может быть установлен из формул (96) и (97) для χ и τ. Для решений уравнений Эйнштейна с Λ-членом они имеют вид

$$\chi = \frac{1}{32\pi^2}\int \left(C_{abcd}C^{abcd} + 2\frac{2}{3}\Lambda^2\right)(g)^{1/2}d^4x, \tag{115}$$

$$\tau = \frac{1}{48\pi^2}\int C_{abcd}{}^*C^{abcd}(g)^{1/2}d^4x. \tag{116}$$

Из (115) видно, что метрика может быть решением лишь при $\chi > 0$. Однако для односвязных многообразий это всегда верно, так как для них $\chi = 2 + B_2$, где B_2— второе число Бетти. Комбинируя (115) и (116), получаем неравенство

$$2\chi - 3|\tau| \geqslant \frac{32c^2}{3}. \tag{117}$$

Из формулы (115) можно видеть, что для больших чисел Эйлера справедливо по крайней мере одно из следующих утверждений:

а) значение c^2 велико,

б) интеграл $\int C_{abcd}C^{abcd}(g)^{1/2}d^4x$ велик.

В первом случае постоянная c должна быть положительной (т. е. Λ должна быть отрицательной), поскольку имеется ограничение на c снизу: $c \geqslant -(^3/_8\lambda)^{1/2}$. Во втором случае должен быть велик тензор Вейля. Как и в обычной общей теории относительности, это приведет к такому же эффекту схождения геодезических, к какому приводит положительный тензор Риччи. Но между любыми двумя точками в пространстве должна быть геодезическая минимальной длины, которая не содержит сопряженных точек. Следовательно, для того, чтобы кривизна Вейля не приводила к слишком быстрому схождению геодезических, нужно ввести отрицательный тензор Риччи и Λ-член порядка $-C_{abcd}C^{abcd}L^2$, где L — некоторый типичный масштаб длины порядка $V^{1/4}\chi^{-1/4}$, т. е. порядка длины, приходящейся на «единицу топологии». Тогда следует ожидать, что два члена в (115) будут сравнимы по величине, а c будет порядка $d\chi^{1/2}$, где $d \leqslant \sqrt{3}/4$.

Это подтверждается рядом примеров, за которые я благодарен Н. Хитчину. Для произведений двумерных многообразий постоянной кривизны имеем $d = 1/4$. Для алгебраических гиперповерхностей $d = \sqrt{2}/8$. Хитчин получил все семейство решений, лежащих между этими пределами. Если, кроме того, решение допускает келерову

структуру, то имеет место равенство

$$3\tau + 2\chi = 32c^2. \qquad (118)$$

Этим результатам можно дать такую интерпретацию: мы имеем набор «гравитационных инстантонов», число которых порядка χ, а действие каждого порядка L^2, где L — характерный размер порядка $V^{1/4}\chi^{-1/4}$. Мы должны также оценить зависимость однопетлевой кривой Z_g от Λ и χ. Зависимость от Λ получается из масштабного поведения и имеет вид

$$Z_g \sim \Lambda^{-\gamma},$$

где

$$\gamma = \int \left(\frac{53}{720\pi^2} C_{abcd}C^{abcd} + \frac{763}{540\pi^2} \Lambda^2 \right) (g_0)^{1/2} d^4x. \qquad (119)$$

Постоянную γ можно считать числом добавочных мод, обусловленных возмущениями фоновой метрики, превышающими уровень возмущений для плоского пространства. Из формулы (119) можно увидеть, что γ того же порядка, что и χ. Поэтому каждому «инстантону» можно приписать некоторое число добавочных мод.

Все вышеизложенное, по-видимому, дает основания для следующей оценки:

$$Z[\Lambda] = \left(\frac{\Lambda}{\Lambda_0} \right)^{-\gamma} \exp(b\chi\Lambda^{-1}), \qquad (120)$$

где $b = 8\pi d^2$, а Λ_0 связано с нормировочной постоянной μ. Учитывая (120) в выражении (113), можно точно вычислить контурный интеграл и получить

$$N(V) = \Lambda_0^{\gamma} \left(\frac{8\pi b\chi}{V} \right)^{1-\gamma/2} I_{\gamma-1} \left(\frac{Vb\chi}{2\pi} \right), \qquad (121)$$

где $V \geqslant 0$.

Однако качественная зависимость от параметров видна более ясно при приближенном вычислении (113) методом стационарной фазы. Действительно, производить более точное вычисление выражения (113) неразумно, так как $Z[\Lambda]$ мы вычисляли лишь в приближении стационарной фазы. Точка стационарности фазы определяется равенством

$$\Lambda_s = 4\pi \frac{\gamma \pm (\gamma^2 + Vb\chi/2\pi)^{1/2}}{V}. \qquad (122)$$

Поскольку контур должен проходить справа от сингулярности при $\Lambda = 0$, перед квадратным корнем следует взять знак плюс.

Значение Λ, соответствующее стационарной фазе, всегда положительно, несмотря на то, что вычисление $Z[\Lambda]$ производилось с использованием фоновых метрик, для которых Λ были отрицательны при больших числах Эйлера. Это означает, что мы должны аналитически продолжить Z от отрицательных значений Λ к положительным. Это аналитическое продолжение эквивалентно умножению

метрики на чисто мнимый конформный множитель, который так или иначе нужен для того, чтобы сделать интеграл по траекториям сходящимся по конформным множителям.

Из приближения стационарной фазы имеем

$$N(V)=Q(\Lambda_s)\equiv\left(\frac{\Lambda_s}{\Lambda_0}\right)^{-\gamma}\exp\left(b\chi\Lambda_s^{-1}+\frac{V\Lambda_s}{8\pi}\right). \tag{123}$$

Главный вклад в $N(V)$ дают топологии, для которых $dQ/d\chi=0$. Если предположить, что $\gamma=a\chi$, где $a=\text{const}$, то получим, что это условие удовлетворяется при

$$-a\ln\left(\frac{\Lambda_s}{\Lambda_0}\right)+b\Lambda_s^{-1}=0. \tag{124}$$

Если $\Lambda_0\gg 1$, этому равенству удовлетворяет $\Lambda_s\approx\Lambda_0$. Если $\Lambda_0<1$, то $\Lambda_s\approx\Lambda_0^{a/b}$. Из уравнения (122) тогда следует, что $\chi=hV$, где коэффициент пропорциональности h зависит от Λ_0. Другими словами, главный вклад в $N(V)$ дают метрики, в которых один гравитационный инстантон приходится на объем h^{-1}.

К каким наблюдаемым эффектам приводит эта пеноподобная структура, еще предстоит выяснить, но в их число, возможно, входят гравитационные распады барионов или мюонов, вызванные их падением на гравитационные инстантоны или виртуальные черные дыры и последующим появлением в виде других частиц. Следует также ожидать несохранения аксиально-векторного тока, вызванного топологиями с неисчезающей сигнатурой τ.

ЛИТЕРАТУРА

1. *Березин Ф. А.* Метод вторичного квантования.— М.: Наука, 1965.
2. *Brill D. R.*, Ann. Phys. (N. Y.), **7**, 46 (1959).
3. *Brill D. R., Deser S.*, Ann. Phys. (N. Y.), **50**, 548 (1968).
4. *Davis L. R.*, неопубликованный доклад, 1978.
5. *Deser S., Zumino B.*, Phys. Lett., **63B**, 335 (1976).
6. *DeWitt B. S.*, In: Relativity Groups and Topology, eds. B. S. and C. DeWitt. Gordon and Breach, New York, 1963.
7. *DeWitt B. S.*, Phys. Rev., **162**, 1195 (1967).
8. *DeWitt B. S.*, Phys. Rep., **196**, 297 (1975). [Имеется перевод: Черные дыры.— М.: Мир, 1978, с. 66.]
9. *Dowker J. S., Critchley R.*, Phys. Rev., **D13**, 3224 (1976).
10. *Eguchi T., Freund P. G. O.*, Phys. Rev. Lett., **37**, 1251 (1976).
11. *Fade'ev L. D., Popov V. N.*, Phys. Lett., **25B**, 697 (1967).
12. *Ferrara S., van Nieuwenhuizen P.*, Phys. Rev. Lett., **37**, 669 (1976).
13. *Feynman R. P.*, In: Magic Without Magic, ed. J. Klauder. W. H. Freemen, San Francisco, 1972.
14. *Freedman D. Z., van Nieuwenhuizen P., Ferrara S.*, Phys. Rev., **D13**, 3214 (1976).
15. *Geroch R. P.*, Ann. N. Y. Acad. Sci., **224**, 108 (1973).
16. *Gibbons G. W.*, Phys. Lett., **60A**, 385 (1977).
17. *Gibbons G. W.*, Phys. Lett., **61A**, 3 (1977).
18. *Gibbons G. W., Hawking S. W.*, Phys. Rev., **D15**, 2752 (1977). [Имеется перевод: Черные дыры.— М.: Мир, 1978, с. 310.]
19. *Gibbons G. W., Hawking S. W.*, Phys. Rev., **D15**, 2738 (1977).
20. *Gibbons G. W., Perry M. J.*, Proc. R. Soc. Lond., **A358**, 467 (1978).

21. *Gibbons G. W., Perry M. J.*, Nucl. Phys., **B146**, 90 (1978).
22. *Gibbons G. W., Pope C. N.*, Commun. Math. Phys., **61**, 239 (1978).
23. *Gibbons G. W., Hawking S. W., Perry M. J.*, Nucl. Phys., **B138**, 141 (1978).
24. *Gilkey P. B.*, The Index Theorem and the Heat Equation. Publish or Perish, Boston, 1974.
25. *Gilkey P. B.*, J. Differ. Geom., **10**, 601 (1975).
26. *Hartle J. B., Hawking S. W.*, Commun. Math. Phys., **26**, 87 (1972).
27. *Hartle J. B., Hawking S. W.*, Phys. Rev., **D13**, 2188 (1976). [Имеется перевод: Черные дыры.— М.: Мир, 1978, с. 222.]
28. *Hawking S. W.*, Phys. Rev., **D13**, 191 (1976).
29. *Hawking S. W.*, Phys. Lett., **60A** 81 (1977).
30. *Hawking S. W.*, Commun. Math. Phys., **55**, 133 (1977).
31. *Hawking S. W., Ellis G. F. R.*, The Large Scale Structure of Space-Time. Cambridge University Press, 1973. [Имеется перевод: *Хокинг С., Эллис Дж.* Крупномасштабная структура пространства-времени.— М.: Мир, 1977.]
32. *Hawking S. W., Pope C. N.*, Phys. Lett., **73B**, 42 (1978).
33. *Israel W., Wilson G. A.*, J. Math. Phys., **13**, 865 (1972).
34. *Jang Pong Soo, Wald R. M.*, J. Math. Phys., **18**, 41 (1977).
35. *Lapedes A.*, в печати.
36. *McKean M. P., Singer I. M.*, J. Differ. Geom., **1**, 43 (1967).
37. *Newman E. T., Unti T., Tamburino L.*, J. Math. Phys., **4**, 915 (1963).
38. *Page D. N.*, Phys. Rev., **D18**, 2733 (1978).
39. *Perry M. J.*, Nucl. Phys. **B**, в печати.
40. *Regge T.*, Nuovo Cimento, **19**, 558 (1961).
41. *Steenrod N.*, The Topology of Fibre Bundles. Princeton University Press, 1951. [Имеется перевод: *Стинрод Н.* Топология косых произведений.— М.: ИЛ, 1953.]
42. *Stelle K.*, Phys. Rev., **D16**, 953 (1977).
43. *Stelle K.*, в печати, 1978.
44. *t'Hooft G., Veltman M.*, Nucl. Phys., **B44**, 189 (1972).
45. *Wheeler J.*, In: Relativity Groups and Topology, eds. B. S. and C. M. DeWitt. Gordon and Breach, New York, 1963.
46. *Yau S. T.*, Proc. Nat. Acad. Sci. USA, **74**, 1798 (1977).
47. *Зельдович Я. Б., Старобинский А.*, ЖЭТФ, **6**, 2161 (1972).
48. *Carter B.*, In: General Relativity. An Einstein Sentenary Survey, Cambridge University Press, Cambridge, p. 294, 1979.
49. *Gibbons G. W.*, там же р. 639.

VIII. УЛЬТРАФИОЛЕТОВЫЕ РАСХОДИМОСТИ В КВАНТОВЫХ ТЕОРИЯХ ГРАВИТАЦИИ

С. Вейнберг [1])

1. ВВЕДЕНИЕ

С тех пор как физики задумались над квантовой теорией гравитации, стало ясно, что ультрафиолетовые расходимости представляют трудную проблему. Простой размерный анализ [2]) указывает, что если константа связи полевой теории имеет размерность [масса]^d (принимая $\hbar = c = 1$), то интеграл для диаграммы Фейнмана порядка N будет вести себя при больших моментах как $\int p^{A-Nd}dp$, где A зависит от рассматриваемого процесса, но не от N. Следовательно, опасными взаимодействиями являются взаимодействия с $d<0$; в этом случае интегралы для любого процесса будут расходиться в довольно высоком порядке. Однако гравитационная постоянная Ньютона имеет размерность $d=-2$ (для $\hbar = c = 1$, $G = 6{,}7 \cdot 10^{-39}$ ГэВ$^{-1}$), поэтому общая теория относительности является прекрасным примером теории с опасными взаимодействиями.

Этот интуитивный аргумент становится точным с развитием ковариантных правил вычисления фейнмановских диаграмм в квантовой теории гравитации [4—10] [3]). Знакомство с этими правилами непосредственно показало, что общая теория относительности не удовлетворяет обычным тестам на перенормируемость. Это было подтверждено детальными расчетами [11—15] [4]): однопетлевой интеграл для вакуумных флуктуаций в классическом фоновом поле

---

[1]) *S. Weinberg*, Department of Physics, Harvard University, Cambridge, Massachusetts, USA.

[2]) Задолго до создания современной теории перенормировок Гейзенберг [1] предложил классификацию взаимодействий элементарных частиц на взаимодействия с безразмерными связями и взаимодействия, связи которых имеют размерность отрицательной степени массы, и он предположил, что характерная масса, которая входит в эти последние связи, будет устанавливать границу применимости существующих теорий. Гейзенберг отметил также, что в фермиевской теории β-распада константа связи имеет размерность (масса)$^{-2}$, и предположил, что динамические эффекты могли быть связаны с энергиями порядка $G_F^{-1/2}$, как, например, в космических ливнях [2]. После создания теории перенормировок было отмечено, что неперенормируемые теории являются именно теми теориями, связи которых имеют размерность отрицательных степеней массы, и что скорости реакций в таких теориях могут быстро расти с энергией [3].

[3]) Вывод ковариантных правил из канонического формализма см. в работах [9, 10].

[4]) Инфракрасные расходимости здесь не ставят никаких проблем (см. работу [16]).

$g_{\mu\nu}$, как выяснилось, имеет расходящиеся члены [1]), пропорциональные R^2 и $R^{\mu\nu}R_{\mu\nu}$. (Эти члены, конечно, обращаются в нуль, если $g_{\mu\nu}$ удовлетворяет вакуумным уравнениям Эйнштейна, $R_{\mu\nu}=0$, но не в присутствии материальных полей с $T_{\mu\nu}\neq 0$.) В более общем случае из размерных соображений следовало бы ожидать, что диаграммы с L витками должны иметь расходящиеся члены, пропорциональные $(L+1)$-й степени тензора кривизны. Как во всякой неперенормируемой теории, сокращение этих ультрафиолетовых расходимостей потребовало бы введения в лагранжиан бесконечного числа членов, пропорциональных произвольно высоким степеням тензора кривизны и его ковариантных производных, но если бы эти члены были введены, расходимости появились бы даже раньше.

Конечно, при решении большинства физических проблем можно прекрасно обойтись без понимания квантовой гравитации и ее ультрафиолетовых расходимостей. Однако глубоко не безразлично, что слияние двух таких фундаментальных теорий, как квантовая механика и общая теория относительности, должно приводить к явному противоречию. Кроме того, попытка построить единую калибровочную теорию слабых, электромагнитных *и* сильных взаимодействий может вынудить нас перейти к рассмотрению более высоких энергий, чем энергии 10^{19} ГэВ [18], при которых гравитация является сильным взаимодействием. Если это так, то успеха в развитии такой «сверхединой» теории придется ждать до тех пор, пока мы не поймем, как обращаться с ультрафиолетовыми расходимостями в квантовой гравитации.

Эта проблема, возможно, возникает потому, что обычный формализм квантовой теории поля, развитый в плоском пространстве, просто не может быть приложен к гравитации. В конце концов, гравитация — это очень специфическое явление, включающее в себя саму топологию пространства и времени. Вероятно также, что все же может быть найден способ описания гравитации вместе с соответствующими материальными полями при помощи обычной перенормируемой квантовой теории поля. Обзор некоторых попыток найти такой способ дан в разд. 2. Однако в данной статье будет главным образом рассматриваться другая возможность, а именно что квантовая теория поля, которая включает гравитацию, может удовлетворять обобщенному варианту условия перенормируемости, известному как *асимптотическая безопасность* [2]).

Говорят, что теория является асимптотически безопасной, если «существенные» параметры связи достигают фиксированной точки, когда масштаб импульса их точки перенормировки стремится к

---

[1]) Необходимость в контрчленах, пропорциональных $R_{\mu\nu}R^{\mu\nu}$ и R^2, была указана очень рано Утиямой и Де Виттом в работе [17].

[2]) Теория, удовлетворяющая условию асимптотической безопасности, называется также «калибровочной теорией с ультрафиолетовой фиксированной точкой». Здесь будет использоваться терминология автора.— *Прим. перев.*

бесконечности. Это условие вводится в разд. 3 как средство избежать нефизических особенностей при очень высокой энергии. Из наблюдаемых свойств фазовых переходов второго рода можно заключить, что для некоторых, а может быть, и всех фиксированных точек условие асимптотической безопасности действует весьма аналогично условию перенормируемости, фиксируя все, кроме, может быть, конечного числа существенных параметров связи теории. Для некоторых фиксированных точек асимптотическая безопасность даже требует, чтобы теория была перенормируемой в обычном смысле.

В разд. 4 рассматривается поведение асимптотически безопасной теории при обычных энергиях, гораздо ниже планковской массы [1]) $M_P=G^{-1/2}=1{,}2\cdot 10^{19}$ ГэВ. Оказывается, что такая теория будет эффективно перенормируемой в обычном смысле, причем все неперенормируемые взаимодействия подавляются как степени E/M_P. Единственное исключение составляет гравитация, которая хотя и является невероятно слабым взаимодействием, тем не менее не может не обнаруживать себя, так как ее когерентность и большой радиус действия позволяют нам наблюдать ее макроскопические проявления; гравитация на расстояниях, гораздо больших, чем планковская длина волны, $M_P^{-1}=1{,}6\cdot 10^{-33}$ см, будет описываться эйнштейновским лагранжианом [2]) $\sqrt{g}R/16\pi G$. Таким образом, если исходить из этой точки зрения, то вполне естественно, что слабые, электромагнитные и сильные взаимодействия микрофизики должны так хорошо описываться перенормируемыми квантово-полевыми теориями и что только одна сила требует, по-видимому, неперенормируемого описания, единственная сила, наблюдаемая лишь по своим макроскопическим проявлениям,— сила гравитации.

Остается вопрос, существуют ли на самом деле какие-нибудь теории гравитации, которые являются асимптотически безопасными. В разд. 5 мы покажем, как использовать продолжение по размерности, чтобы попытаться ответить на этот вопрос. Этот метод применяется к гравитации в разд. 6, и он приводит к выводу, что имеется асимптотически безопасная теория чистой гравитации в $(2+\varepsilon)$ измерениях (здесь $0<\varepsilon\ll 1$). Материальные поля могут менять или не менять этот вывод в зависимости от их типа и числа. К сожалению, проблема продолжения от $2+\varepsilon$ к 4 измерениям остается нерешенной.

Как читатель вскоре убедится, представленный здесь анализ чрезвычайно «сырой»: многое предполагается, но мало доказывается фактически относительно квантовой гравитации. Я пользуюсь случаем, предоставленным этим изданием, скорее для того, чтобы высказать свою собственную точку зрения о наиболее вероятном

---

[1]) Планковская масса дается (для $\hbar=c=1$) в терминах ньютоновской постоянной G величиной $G^{-1/2}=1{,}2.10^{19}$ ГэВ (см. [19]).

[2]) В обозначениях, используемых в этой статье, $\hbar=c=1$; метрика плоского пространства $\eta_{\mu\nu}$ имеет диагональные элементы $+$, $+$, $+$, $-$, знаковые условия для R, $R_{\mu\nu}$ и т. д. такие же, как в работе [20].

направлении прогресса в разрешении проблемы ультрафиолетовых расходимостей в квантовой теории гравитации, чем продемонстрировать законченный формализм.

Другая цель, которую я здесь преследовал, состояла в том, чтобы дать обзор некоторых математических методов, разработанных в теориях критических явлений и элементарных частиц, которые выступают под общей рубрикой «ренормализационной группы» (ренорм-группы) [1]. Очень мало из этого материала является новым и почти ничего не принадлежит мне, но я надеюсь, что, собранный мною в этой статье, он даст некоторым читателям опыт обращения с ренормализационной группой, который мог бы оказаться полезным в будущей работе в области квантовой теории гравитации.

2. ПЕРЕНОРМИРУЕМЫЕ ТЕОРИИ ГРАВИТАЦИИ

Насколько я знаю, имеется ровно три подхода, с помощью которых можно было бы надеяться построить перенормируемую теорию гравитации.

а. Расширенные теории гравитации

Могло случиться, что в результате объединения гравитации с соответствующими материальными полями и наложения соответствующих симметрий все ультрафиолетовые расходимости оказались бы сокращенными, за исключением конечного их числа, которые могли бы быть поглощены при перенормировке параметров теории. В этом случае расширенная теория материи и гравитации была бы перенормируемой в обычном смысле.

Как отмечалось в разд. 1, анализ размерностей наводит на мысль, что квантово-полевая теория гравитации будет всегда иметь ультрафиолетовые расходимости. Опыт вообще показывает, что если ультрафиолетовая расходимость ожидается из размерных соображений и не исключается какой-нибудь симметрией теории, то ультрафиолетовая расходимость действительно будет существовать. Вопрос, исключается ли ультрафиолетовая расходимость симметрией, эквивалентен вопросу о том, запрещает ли симметрия существование в лагранжиане члена, который требовался бы, чтобы дать контрчлен для этой расходимости. Следовательно, по-видимому, единственная надежда для расширенной теории гравитации и материи, перенормируемой в обычном смысле, заключается в том, что симметрии теории исключат из лагранжиана все (кроме, может быть, конечного числа) типы взаимодействий.

---

[1]) Идея ренорм-группы была введена в ее современном виде в физику элементарных частиц Гелл-Манном и Лоу [21]. Похожие идеи обсуждались также Штюкельбергом и Петерманом [22]. Относительно приложений к теории критических явлений см. работы [23 — 28].

Это, конечно, возможно для обычных симметрий. Если симметрии допускают построение набора инвариантных взаимодействий $\sqrt{g}\mathscr{L}_i$, то мы можем построить бесконечное число других дозволенных взаимодействий, образуя произведения $\sqrt{g}\mathscr{L}_i\mathscr{L}_j$, $\sqrt{g}\mathscr{L}_i\mathscr{L}_j\mathscr{L}_k$ и т. д., и можно ожидать появления любой из бесконечного числа соответствующих ультрафиолетовых расходимостей.

Однако возможно, что симметрии теории не допускают построение *любого* инвариантного взаимодействия, а допускают только такие взаимодействия $\sqrt{g}\mathscr{L}_i$, вариации которых $\delta\sqrt{g}\mathscr{L}_i$ под действием преобразований группы симметрии являются полными производными. Тогда вклады этих взаимодействий $\int d^4x\sqrt{g}\mathscr{L}_i$ в действие инвариантны, поэтому они являются допустимыми взаимодействиями. Однако вариации в произведениях $\sqrt{g}\mathscr{L}_i\mathscr{L}_j$, $\sqrt{g}\mathscr{L}_i\mathscr{L}_j\mathscr{L}_k$ и т. д. не являются, вообще говоря, полными производными, поэтому эти взаимодействия могут оказаться запрещенными симметриями теории, и соответствующие ультрафиолетовые расходимости не будут появляться.

Наиболее интересные примеры теории этого типа даются теориями суперсимметрии [1]), в которых частицы с различными спинами соединяются в неприводимых представлениях. Лагранжиан не является фактически инвариантом относительно преобразований суперсимметрии, а служит отдельным членом супермультиплета, который преобразуется в полную производную. Суперсимметрия налагает такие жесткие условия на лагранжиан, что в некоторых теориях «супергравитации» [2]) [10] не имеется никаких допустимых взаимодействий такого вида, который был бы необходим для сокращения расходимостей в диаграммах с одной или двумя петлями [3]); в ре-

---

[1]) Суперсимметрия была введена Вессом и Зумино [29 — 31]. Аналогичные идеи разрабатывались также Гольфандом и Лихтманом [32] и Волковым и Акуловым [33]. Формулировка «суперпространства» была дана в работе [34] (обзор см. [35]).

[2]) «Супергравитация» является суперсимметричной теорией мультиплета, состоящего из гравитона плюс безмассовой майорановской частицы спина $^3/_2$. Она была введена в работах [36, 37]. Обзоры с обсуждениями теорий «расширенной супергравитации» см. в работах [38 — 40]. Относительно теорий супергравитации, основанных на обобщениях лагранжиана $\sqrt{g}R^2$, см. [41, 42].

[3]) Сокращение некоторых расходимостей в глобально-суперсимметричных полевых теориях материи было отмечено в неопубликованной работе Ли (цитируется в [43—45]). Сокращение расходимостей в гравитационных функциях Грина в теории суперсимметричных материальных полей было отмечено Зумино [46]. Это позволяет предположить, что суперсимметричная теория гравитации могла бы быть перенормируемой (см. [47]). Сокращение однопетлевых расходимостей в теории чистой супергравитации было отмечено в работах [48—51]. Однопетлевые расходимости в теориях с супергравитацией, связанной с материей, были найдены в точных вычислениях [52] и объяснены в общем виде в работе [53]. Однако конечность однопетлевых графиков для $O(n)$-расширенной супергравитации была показана в работе [54] для $n=2$, в [55] — для $n=3$ и $n=4$ и в [56] — для $n=8$. В двухпетлевом порядке было показано, что чистая супергравитация свободна от

зультате в этих теориях не встречается никаких ультрафиолетовых расходимостей в низших порядках теории возмущений. Однако пока нет оснований ожидать, что эта теория не будет содержать бесконечного числа допустимых взаимодействий, которые нужны в качестве контрчленов в высших порядках теории возмущений.

б. Пересуммирование

Сохранилась надежда, что неперенормируемая теория могла бы быть сделана эффективно перенормируемой перераспределением членов ряда теории возмущений [1]). Некоторые бесконечные подгруппы графиков были бы просуммированы в первую очередь, и если бы эти суммы достаточно быстро стремились к нулю при больших моментах, они могли бы использоваться в качестве строительных блоков в новом наборе фейнмановских правил с лучшей сходимостью. Однако трудно показать, что такое изменение порядка суммирования ведет к более надежному ряду теории возмущений, или в действительности найти ему оправдание в каких-нибудь соображениях, отличных от соображений, связанных с исключением ультрафиолетовых расходимостей. На самом деле, во многих случаях мы можем непосредственно видеть, что члены конечного порядка содержатся в пересуммированном ряду возмущений, и это, повидимому, относится к нефизическим особенностям.

Наиболее простой пример такого типа пересуммирования в квантовой теории гравитации дает теория, основанная на лагранжиане [2])

$$\mathscr{L} = -\frac{1}{16\pi G}\sqrt{g}\,R - f\sqrt{g}\,R^2 - f'\sqrt{g}\,R_{\mu\nu}R^{\mu\nu}. \qquad (1)$$

Действуя обычным образом, нужно было бы рассматривать два последних члена в качестве возмущения и разлагать в ряды по степеням f, f', а также G. На этом пути, конечно, встретились бы ужасные ультрафиолетовые расходимости, но не было бы нефизических особенностей в любом конечном порядке теории возмущений. Однако

ультрафиолетовых расходимостей [57 — 59]. То же самое оказалось верным в одно- и двухпетлевом порядках для $O(n)$-расширенной супергравитации с n генераторами суперсимметрии. Однако в цитированной выше работе Дезера, Кэя и Стелле и в работе [60] было указано, что имеются возможные ультрафиолетовые расходимости в чистой супергравитации в трехпетлевом порядке, которые уже не могут быть исключены по соображениям симметрии. Аналогичные проблемы были обнаружены в трехпетлевом порядке в $O(2)$-расширенной супергравитации (Deser S., Kay J. H., частное сообщение). Возможные ультрафиолетовые расходимости в супергравитации были проанализированы для всех порядков теории возмущений в работе [61]. Возможные расходимости найдены в каждом порядке выше двух петель, но еще неизвестно, являются ли какие-нибудь из их коэффициентов ненулевыми.

[1]) Многообразные методы изменения порядка суммирования в общей теории относительности или в аналогичных теориях были изучены в работах [62—71].

[2]) Перенормируемые теории гравитации, основанные на лагранжиане, который включает члены $\sqrt{g}R^2$ и $\sqrt{g}R^{\mu\nu}R_{\mu\nu}$, были предложены Дезером [72] и Вейнбергом [73]. Общее исследование таких теорий было проведено Стелле [74].

мы можем просуммировать вклады в R^2 и $R^{\mu\nu}R_{\mu\nu}$ членов, которые квадратичны по гравитационному полю $h_{\mu\nu}$ (где, как обычно, $g_{\mu\nu}=\eta_{\mu\nu}+\sqrt{32\pi G}h_{\mu\nu}$). Это дает модифицированный гравитационный пропагатор, в котором обычный член $1/k^2$ заменяется на

$$\frac{1}{k^2}+\frac{1}{k^2}\alpha fGk^4\frac{1}{k^2}+\frac{1}{k^2}\alpha fGk^4\frac{1}{k^2}\alpha fGk^4\frac{1}{k^2}+\ldots=(k^2-\alpha fGk^4)^{-1}, \qquad (2)$$

где α — безразмерная функция f'/f. (В качестве альтернативы мы могли бы просто рассматривать квадратичную часть членов R^2 и $R_{\mu\nu}R^{\mu\nu}$ как часть лагранжиана нулевого порядка и получить (2) непосредственно как «голый» пропагатор для этой теории). Он достаточно быстро стремится к нулю на бесконечности, так что лагранжиан (1) теперь проходит обычные степенные тесты на перенормируемость. Неудобно, конечно, что (2) имеет теперь полюс при $k^2=-1/\alpha fG$ с вычетом $-1/\alpha^2f^2G^2$ неподходящего знака, чтобы согласовываться с унитарностью [1]).

Встретившаяся здесь проблема является частным случаем более общей трудности. При обычных предположениях аналитичности любое изменение порядка суммирования будет вообще заменять $1/k^2$ в гравитационном пропагаторе на интеграл [76, 77] [2])

$$\int\frac{\rho(\mu)}{\mu^2+k^2}d\mu^2. \qquad (3)$$

Унитарность требует $\rho(\mu)\geqslant 0$ (за исключением, возможно, вклада петлевых графиков, содержащих гравитоны или ду́хи Фаддеева — Попова в лоренц-ковариантных калибровках). Пока $\rho(\mu)\geqslant 0$, интеграл (3) не может убывать быстрее, чем $1/k^2$ при $k^2\to\infty$.

Мы можем увидеть это снова при изменении порядка суммирования, основанном на «$1/N$»-приближении [3]). Предположим, что гравитация связана с N типами материальных полей с гравитационной константой G порядка $1/N$. Для больших N доминирующие диаграммы для полного пропагатора состоят из цепочки петель типа собственной энергии, каждая цепочка построена из одной петли линий материи и связана с голым гравитонным пропагатором. (Два множителя $1/\sqrt{N}$, вводимые связыванием гравитона с каждой петлей материи, сокращаются с множителем N от суммы всех типов полей в петле материи, поэтому эти диаграммы имеют нулевой по-

[1]) Изменение порядка суммирования, которое могло бы исключить этот полюс, было предложено в работе [75]. Аналогичная идея высказывалась Дезером (Deser S., частное сообщение, 1978).

[2]) Вычитания могли бы понадобиться в дисперсионном соотношении для пропагатора, но это сделало бы его асимптотическое поведение еще хуже, чем (3).

[3]) Приближение «большого N» было развито в статистической физике в работах [78—81] и в релятивистской квантовой теории поля в работах [82—86]. Нефизические «тахионные» полюсы были найдены в некоторых теориях в работах [87, 88]. Однако было указано, что эти полюсы отсутствуют в устойчивых решениях теории [89]. Фиксированные точки в неперенормируемых теориях поля были исследованы в приближении «большого N» в работе [90].

рядок по $1/N$). Материально-петлевые графики логарифмически расходятся, поэтому, чтобы сократить эти расходимости, мы должны добавить в лагранжиан члены, пропорциональные $\sqrt{g}R^2$ и $\sqrt{g}R^{\mu\nu}R_{\mu\nu}$, и квадратичные части этих членов должны быть просуммированы вместе с петлевыми членами для материи во всех порядках. Когда это сделано, член $1/k^2$ в гравитонном пропагаторе оказывается замененным на [91]

$$k^{-2} \to [k^2 + GNk^4\alpha \ln(k^2/\mu^2)]^{-1}, \quad (4)$$

где α — безразмерная постоянная и μ — точка перенормировки, которая определяет значения f и f'. Пропагатор убывает достаточно быстро при $k \to \infty$, так что не появляется никаких новых ультрафиолетовых расходимостей в графиках высшего порядка по $1/N$. Однако пропагатор (4) теперь имеет особенность при комплексных значениях k^2. Это, несомненно, нарушает наши обычные представления об аналитичности, хотя не ясно, до какой степени комплексные особенности такого сорта неизбежно ведут к конфликту с фундаментальными физическими принципами [92].

в. Составные гравитоны

Не существует никакой перенормируемой теории 3—3-резонансного поля, но это никого не беспокоит: 3—3-резонанс считается трехкварковым связанным состоянием, возникающим в перенормируемой теории сильного взаимодействия, известной как квантовая хромодинамика. Аналогичным образом гравитон может быть составной частицей нулевой массы и спина два, который возникает в некоторой перенормируемой квантовой теории поля.

Если гравитон есть просто связанное состояние, то почему он должен описываться теорией, которая является столь изысканно геометрической, как общая теория относительности? Возможный ответ может быть найден в рамках специальной теории относительности и квантовой механики в плоском пространстве. Очень трудно включить безмассовые частицы спина два в лоренц-инвариантную квантовую теорию дальнодействующих сил; необходимо связать такую частицу с сохраняющимся тензором энергии-импульса, который включает сами гравитоны. Фактически было показано [93] [1]), что безмассовые частицы спина два должны описываться эффективной полевой теорией, удовлетворяющей принципу эквивалентности. В той степени, в которой мы интересуемся только проявления-

---

[1]) Программа выведения классической общей теории относительности из квантовой механики и специальной теории относительности была завершена в работе [94]. Я слышал, что подобные идеи развивались Р. Фейнманом в неопубликованных лекциях в Калифорнийском технологическом институте.

ми гравитации на больших расстояниях, единственная такая теория — это общая теория относительности [1]).

Разъясняющий пример таких эффективных полевых теорий дается теорией мягких пионов, известной как киральная динамика [2]). В пределе стремящихся к нулю масс *u*- и *d*-кварков из квантовой хромодинамики следует, что сильные взаимодействия обладают глобальной симметрией, киральной $SU(2)\otimes SU(2)$. Эта симметрия спонтанно нарушается, давая безмассовый бозон Голдстоуна — пион. Даже если пион состоит из пары кварк — антикварк, его взаимодействия описываются эффективной полевой теорией [99—103], в которой пион представлен полем $\vec{\pi}$, преобразующимся согласно нелинейной трехмерной реализации $SU(2)\otimes SU(2)$. При одном удобном определении пионного поля лагранжиан киральной динамики принимает вид

$$\mathcal{L} = -\frac{1}{2F_\pi^2} D_\mu\vec{\pi}\cdot D^\mu\vec{\pi} - f\,(D_\mu\vec{\pi}\cdot D^\mu\vec{\pi})^2 \ldots, \tag{5}$$

где $D_\mu\vec{\pi}$ — «ковариантная производная»,

$$D_\mu\vec{\pi} = \partial_\mu\vec{\pi}/(1+\vec{\pi}^2), \tag{6}$$

причем $F_\pi \approx 190$ МэВ и $f = O(1)$ — постоянные, значения которых должны быть взяты из эксперимента. Предположим, что мы используем эту теорию, чтобы вычислить какую-нибудь из инвариантных амплитуд $M(E)$ для пион-пионного рассеяния при фиксированном угле и энергии E. Для достаточно малых E матричный элемент дается членом первого порядка в квадратичной части первого члена в (5), поэтому $M(E) \propto E^2/F_\pi^2$. В следующем порядке по E мы должны учесть однопетлевые графики, построенные из первого члена в (5), и эффекты первого порядка от второго члена в (5). Однопетлевые графики, конечно, расходятся, но расходимость может быть включена в перенормировку f, давая в итоге матричный элемент.

$$M(E) = \frac{aE^2}{F_\pi^2} + \frac{bE^4}{F_\pi^4}\ln\left(\frac{E}{\mu}\right) + cf(\mu)\frac{E^4}{F_\pi^4} + O(E^6\ln^2 E). \tag{7}$$

Здесь a, b и c — известные безразмерные величины, которые зависят от переменных угла и изоспина; $f(\mu)$ — перенормированное значение f; μ — единица массы, используемая в определении $f(\mu)$. Даже если эффективная теория является неперенормируемой и в принципе включает бесконечное число неизвестных параметров,

---

[1]) Эйнштейн вывел свои полевые уравнения как единственные общековариантные уравнения, в которых каждый член (слева) включал бы именно две производные метрики. Общая ковариантность является как раз подходящим способом обеспечить выполнение принципа эквивалентности, а ограничение двумя производными позволяет выбрать те члены в самых общих общековариантных полевых уравнениях, которые существенны на больших расстояниях. Обсуждение полевых уравнений Эйнштейна в этом аспекте см. в работе [95].

[2]) См. обзоры [96—98].

мы можем вычислить члены E^2 и $E^4 \ln E$ в пион-пионном рассеянии в терминах единственной константы F_π.

Почти таким же образом независимо от того, является ли гравитон связанной частицей, требования лоренц-инвариантности и квантовой механики заставляют эффективный гравитационный лагранжиан принять вид [93]

$$\mathscr{L}_G = -\frac{1}{16\pi G}\sqrt{g}\,R - f\sqrt{g}\,R^2 - f'\sqrt{g}\,R_{\mu\nu}R^{\mu\nu} - 16\pi G f''\sqrt{g}\,R^3 - \ldots, \tag{8}$$

а лагранжиан материи должен зависеть от $g_{\mu\nu}$ таким образом, чтобы быть общековариантным. Его можно использовать, чтобы построить ряд теории возмущений по степеням GE^2 или G/r^2 (где E и r — энергия и длина, которые характерны для изучаемого процесса), поэтому в качестве невозмущенного лагранжиана следует принять квадратичную часть одного только первого члена $-\sqrt{g}R/16\pi G$. Поскольку квадратичные части второго и третьего членов должны здесь рассматриваться как возмущения более высокого порядка, мы не встретимся с нефизическими особенностями, обсуждавшимися в разд. 2, *б*. В разд. 4 будет показано, что полный эйнштейновский член $-\sqrt{g}R/16\,\pi G$ может быть использован в древесном приближении, чтобы вычислить ведущие члены во всех многогравитонных функциях Грина, и может даже быть использован в однопетлевом приближении для расчета квантовых поправок относительного порядка $GE^2 \ln E$ или $(G/r^2) \ln r$.

Главное возражение против того, чтобы рассматривать гравитон как составную частицу, состоит просто в том, что не ясно, почему связанное состояние спина два должно иметь точно нулевую массу. Вероятно, это можно было бы понять либо динамически [1]), либо в терминах суперсимметрии (см. сноску 2 на с. 411), включающей гравитон в мультиплет с другими составными частицами, которые должны быть безмассовыми вследствие киральной или иной симметрии. Тем не менее это обсуждение показывает, что если мы хотим лишь изучить низкоэнергетические или дальнодействующие свойства пионов или гравитонов, то нет необходимости знать что-нибудь о механизме, которым связываются эти частицы, или даже о том, являются ли вообще пионы и гравитоны составными частицами. В разд. 4 будет дано несколько более систематическое рассмотрение низкочастотного предела квантовых теорий гравитации, не требующее обсуждения вопроса о том, составными или элементарными частицами являются гравитоны.

---

[1]) Неустойчивости образования фотонных пар были предложены в качестве возможного источника гравитации в работах [104, 105]. Другие теории с составными гравитонами см. в работах [106—115].

3. АСИМПТОТИЧЕСКАЯ БЕЗОПАСНОСТЬ

Допустим в порядке дискуссии, что ни один из подходов, упомянутых в предыдущем разделе, не может быть успешно применен к гравитации. Предположим, что никакое объединение гравитации с материальными полями не приводит к теории, которая является перенормируемой в обычном смысле, и что никакие перегруппировки в суммах не приводят к перенормированной теории возмущений, совместимой с унитарностью, а также что гравитон не может быть интерпретирован как составная частица, возникающая в перенормируемой теории поля, лежащей в ее основе. Если вообще гравитация должна описываться квантово-полевой теорией в плоском пространстве, то мы можем столкнуться с перспективой иметь дело с теорией, которая не является перенормируемой в обычном смысле. В данном разделе будет описываться обобщенный вариант условия перенормируемости, который все же мог бы быть применимым к гравитации в этом случае.

В любой неперенормируемой теории ультрафиолетовые расходимости появляются во всех функциях Грина (и во всех порядках по их внешним импульсам), поэтому, чтобы обеспечить контрчлены для этих бесконечностей, мы должны включить в лагранжиан все возможные взаимодействия, допустимые симметриями теории. Следовательно, гравитационный лагранжиан должен был бы включать не только эйнштейновский член $-\sqrt{g}R/16\pi G$, но также члены, пропорциональные $\sqrt{g}R^2$, $\sqrt{g}R^{\mu}_{\nu}R^{\nu}_{\mu}$, $\sqrt{g}R^3$ и т. д., плюс члены, включающие произвольные степени материальных, а также гравитационных полей.

Можно было бы полагать, что включение членов, пропорциональных $\sqrt{g}R^2$ и $\sqrt{g}R^{\mu}_{\nu}R^{\nu}_{\mu}$, приведет к нефизическим особенностям типа обсуждавшихся в разд. 2, *б*. Однако эти нефизические особенности не встречаются в теории возмущений, пока вклады этих взаимодействий, квадратичные по полям, суммируются во всех порядках, как в (2). Вообще говоря, это частичное суммирование не может быть оправдано — очевидные полюсы появляются для значений моментов порядка планковской массы[1] $G^{-1/2}=1,2\times 10^{19}$ ГэВ, но при таких моментах гравитация является столь сильной, что теория возмущений вообще перестает работать. Лагранжиан с членами $\sqrt{g}R^2$ и $\sqrt{g}R^{\mu}_{\nu}R^{\nu}_{\mu}$ не будет приводить к нефизическим особенностям при энергиях $E\ll 10^{19}$ ГэВ, пока теория возмущений заслуживает доверия, но мы не можем использовать теорию возмущений, чтобы сказать, будут ли существовать нефизические особенности при энергиях порядка 10^{19} ГэВ в такой теории или даже в самóй общей теории относительности. Вопрос о возможных нефизических особенностях при очень высоких энергиях является тем не ме-

---

[1]) См. примечание 1 на с. 409.

нее важным вопросом и фактически служит ключевым свойством, которое приводит нас к требованию асимптотической безопасности.

Чтобы исследовать такие особенности, мы используем метод ренорм-группы [1]). Пусть $g_i(\mu)$ означает полный набор всех перенормированных параметров связи теории, определенных в точке перенормировки при импульсах, характеризуемых энергетическим масштабом μ. Если $g_i(\mu)$ имеет размерность [масса]^{d_i}, заменим ее на безразмерную связь

$$\bar{g}_i(\mu)=\mu^{-d_i}g_i(\mu). \tag{9}$$

Любой тип парциальной или полной скорости реакции R может быть записан в форме

$$R=\mu^D f\left(\frac{E}{\mu},\ X,\ \bar{g}(\mu)\right), \tag{10}$$

где D — обычная размерность R (например, для полного поперечного сечения $D=-2$); E — некоторая энергия, характеризующая процесс; X символизирует все остальные безразмерные физические переменные, включая все отношения энергий. Основная идея метода ренорм-группы сводится к пониманию того, что физическая величина R не может зависеть от произвольного выбора точки перенормировки μ, в которой определяются связи, так что мы можем принять за μ в (10) любое другое значение, и в частности $\mu=E$, при котором (10) принимает вид

$$R=E^D f(1,\ X,\ \bar{g}(E)). \tag{11}$$

Таким образом, не считая тривиального масштабного фактора E^D, высокоэнергетическое поведение скоростей реакций зависит от поведения связей $\bar{g}(\mu)$ при $\mu\to\infty$.

Здесь имеется техническая деталь, которая заслуживает некоторого пояснения. В набор параметров связи $g_i(\mu)$ входят массы частиц $m(\mu)$ с размерностью $d=+1$. Соответствующие безразмерные параметры (9) имеют вид $m(\mu)/\mu$, и разумно было бы ожидать, что они стремятся к нулю [2]) при $\mu\to\infty$. Однако скорости реакций будут вообще содержать особенности при нулевой массе, соответствующие инфракрасным расходимостям, присутствующим в безмассовой квантовой теории поля. Поэтому невозможно оценить высокоэнергетическое поведение произвольных скоростей реакций, просто полагая массы равными нулю. Именно по этой причине метод ренорм-группы исторически был использован для вычисления высокоэнергетического поведения функций Грина вдали от массовой поверхности, где никакие инфракрасные расходимости не встречаются даже для нулевой массы. Однако массовые особенности могут

---

[1]) См. примечание на с. 410.

[2]) Условие того, что $m(\mu)/\mu$ должно обращаться в нуль для $\mu\to\infty$, обсуждается в работе [116].

быть устранены из самих скоростей физических реакций путем выполнения соответствующего суммирования по определенным наборам начальных и конечных состояний [117—119]. В дальнейшем мы будем подразумевать, что это сделано.

Мы делаем здесь акцент на скорости реакций, а не на функции Грина вне массовой поверхности, и это имеет очень важное преимущество. Матричные элементы на массовой поверхности и скорости реакций не зависят от того, как определяются поля, поэтому они являются функциями только «существенных» параметров связи, т. е. тех комбинаций параметров связи в лагранжиане, которые не меняются, когда мы подвергаем поля точечному преобразованию (как, например, $\phi \to \phi + \phi^2$ для скалярного поля φ). Напротив, функции Грина вне массовой поверхности, конечно, отражают определение рассматриваемых полей и будут потому функциями всех параметров связи в лагранжиане, включая те «несущественные» параметры связи (типа перенормировочных констант поля z), которые не являются инвариантами при переопределении полей. Здесь всегда будет подразумеваться, что $g_i(\mu)$ включают только *существенные параметры* связи теории.

(Имеется известный тест, который может быть использован для идентификации несущественных параметров связи в любой теории. Когда мы изменяем какой-нибудь неперенормированный параметр связи γ_0 на бесконечно малую величину ε, весь лагранжиан изменяется на

$$\mathscr{L} \longrightarrow \mathscr{L} + \varepsilon \frac{\partial \mathscr{L}}{\partial \gamma_0}.$$

Предположим, что мы пытаемся осуществить это изменение простым переопределением полей

$$\psi_n(x) \longrightarrow \psi_n(x) + \varepsilon F_n(\psi(x),\ \partial_\mu \psi(x),\ \ldots).$$

Изменение, вносимое при этом в $\mathscr{L}$, имеет вид

$$\delta\mathscr{L} = \varepsilon \sum_n \left[\frac{\partial \mathscr{L}}{\partial \psi_n} F_n + \frac{\partial \mathscr{L}}{\partial(\partial_n \psi_n)} \partial_n F_n + \ldots\right] =$$

$$= \varepsilon \sum_n \left[\frac{\partial \mathscr{L}}{\partial \psi_n} - \partial_\mu \left(\frac{\partial \mathscr{L}}{\partial(\partial_\mu \psi_n)}\right) + \ldots\right] F_n + \text{члены с полными производными}.$$

Таким образом, изменение $\delta\mathscr{L} = \varepsilon \partial\mathscr{L}/\partial\gamma_0$ в лагранжиане может быть осуществлено переопределением полей, если и только если мы можем найти функции полей F_n и их полные производные, такие, что

$$\frac{\partial \mathscr{L}}{\partial \gamma_0} = \sum_n \left[\frac{\partial \mathscr{L}}{\partial \psi_n} - \partial_\mu \frac{\partial \mathscr{L}}{\partial(\partial_\mu \psi_n)} + \ldots\right] F_n(\psi,\ \partial_\mu \psi,\ \ldots) + \text{члены с полными производными}.$$

Другими словами, *параметр связи γ_0 является несущественным, если и только если $\partial\mathscr{L}/\partial\gamma_0$ обращается в нуль или является полной*

производной, когда мы используем уравнение Эйлера — Лагранжа

$$0=\frac{\partial\mathcal{L}}{\partial\psi_n}-\partial_\mu\frac{\partial\mathcal{L}}{\partial(\partial_\mu\psi_n)}+\ldots .$$

Например, в перенормируемой теории скалярного поля с лагранжианом

$$\mathcal{L}=-\frac{1}{2}Z(\partial_\mu\phi\partial^\mu\phi+m^2\phi^2)-\frac{1}{24}\lambda Z^2\phi^4$$

перенормировочная константа поля является несущественной связью, поскольку мы можем написать

$$\frac{\partial\mathcal{L}}{\partial Z}=\frac{1}{2}\phi\left(\Box^2\phi-m^2\phi-\frac{1}{6}\lambda Z\phi^3\right)-\frac{1}{2}\partial_\mu(\phi\partial^\mu\phi),$$

и первый член обращается в нуль, когда мы используем полевое уравнение

$$\Box^2\phi=m^2\phi+\frac{1}{6}\lambda Z\phi^3.$$

С другой стороны, ни масса m, ни связь λ, ни любая комбинация m и λ не являются несущественными. В этом примере переопределение поля $\phi\to\phi+\varepsilon F$, связанное с единственной несущественной связью Z, является простым изменением масштаба с $F\propto\phi$, но это именно потому, что от теории требуется, чтобы она была перенормируемой; более сложные преобразования с F, являющейся нелинейной функцией ϕ и ее производных, порождали бы неперенормируемые члены в $\mathcal{L}$. В неперенормируемых теориях мы должны рассматривать все возможные переопределения полей, совместимые с их свойствами симметрии, и в результате имеется бесконечное число и несущественных и существенных связей.)

Как мы увидим, работая только с существенными связями, можно сформулировать условие асимптотической безопасности в сжатом виде. Кроме того, оно хорошо стыкуется с «методом фонового поля» [120—125], в котором перенормировка параметров связи устанавливается вычислением матричных элементов между «in» и «out» вакуумами в классическом фоновом поле, которое удовлетворяет полевым уравнениям Эйлера — Лагранжа.

Вернемся теперь к проблеме определения поведения существенных связей $\bar{g}_i(\mu)$. Изменение в $\bar{g}_i(\mu)$ при заданном частичном изменении в μ является безразмерной величиной и может поэтому зависеть от всех $\bar{g}_i(\mu)$, а не отдельно от самого μ, поскольку здесь не имеется других размерных параметров, с которыми можно было бы сравнивать μ. Поэтому скорость изменения $\bar{g}_i(\mu)$ относительно изменения масштаба точки перенормировки μ может быть переписана в виде обобщенного уравнения Гелл-Манна — Лоу [1])

$$\mu\frac{d}{d\mu}\bar{g}_i(\mu)=\beta_i(\bar{g}(\mu)). \tag{12}$$

[1]) См. примечание на с. 410.

Каждая отдельная теория характеризуется *траекторией* в пространстве константы связи, порождаемой решением (12) с заданными начальными условиями.

Функция $\beta(\bar{g})$ может быть вычислена как степенной ряд по $\bar{g}_i$, но это, вообще говоря, не поможет нам установить поведение $\bar{g}(\mu)$ при $\mu\to\infty$. Однако мы можем идентифицировать один общий класс теорий, в которых нефизические особенности почти несомненно отсутствуют. Если связи $\bar{g}_i(\mu)$ достигают «фиксированной точки» g^* при $\mu\to\infty$, то (11) дает простое масштабно-инвариантное поведение $R\propto E^D$ для $E\to\infty$. Чтобы $\bar{g}(\mu)$ достигало g_i^* при $\mu\to\infty$, функция $\beta_i(\bar{g})$ должна обращаться в этой точке в нуль

$$\beta_i(g^*)=0, \tag{13}$$

а связи должны лежать на траектории $\bar{g}_i(\mu)$, которая на самом деле попадает в фиксированную точку. Поверхность, образованная такими траекториями, будет называться *ультрафиолетовой критической поверхностью*. Обобщенный вариант перенормируемости, который мы хотим предложить для квантовой теории гравитации, заключается в том, что константы связи должны лежать на ультрафиолетовой критической поверхности некоторой фиксированной точки. Такие теории будут называться *асимптотически безопасными*[1]).

(Между прочим, требование, чтобы параметры связи достигали фиксированной точки при $\mu\to\infty$, не может быть, вообще говоря, удовлетворено, если мы включим вместе с существенными также и несущественные параметры связи. Например, уравнения ренормгруппы не могут изменить своей формы, когда мы умножаем каждое поле на независимую постоянную; следовательно, если константы перенормировки поля $Z_r(\mu)$ удовлетворяют этим уравнениям, то это же выполняется для $Z_r(\mu)$, умноженных на произвольные постоянные, и поэтому уравнения для $Z_r(\mu)$ должны принимать вид

$$\mu\frac{d}{d\mu}Z_r(\mu)=Z_r(\mu)\,\gamma_r(\bar{g}(\mu)).$$

Вообще говоря, нет причины, по которой $\gamma_r(g^*)$ должно было бы обращаться в нуль или расходиться, поэтому решение для $\mu\to\infty$ будет иметь форму

$$Z_r(\mu)\sim\mu^{\gamma_r(g^*)}.$$

Это вводит поправки к скейлингу в функциях Грина вне массовой поверхности. Однако скорости реакций не зависят от Z, поэтому они могут проявлять «наивный» скейлинг $R\to E^D$, даже если $Z_r(\mu)\to\infty$ для $\mu\to\infty$.)

---

[1]) Более ранние обсуждения — в докладе [126]. Близкие идеи рассматривались в работе [127].

Мы не знаем на самом деле, что теория, которая не является асимптотически безопасной, будет иметь нефизические особенности — предположение асимптотической безопасности является как раз одним из способов обрести разумную уверенность в том, что нефизические особенности не появятся. В качестве примера того, что может случиться, когда теория является *не* асимптотически безопасной, рассмотрим образец дифференциального уравнения [1]).

$$\mu \frac{d}{d\mu} \bar{g}_i = a_i \sum_j (\bar{g}_j - g_j^*)^2, \tag{14}$$

где a_i и g_i^* — наборы произвольных постоянных. Траекториями, достигающими фиксированной точки g^*, являются, очевидно, траектории с начальными значениями вдоль линии

$$\bar{g}_i = g_i^* - a_i \xi, \quad \xi > 0, \tag{15}$$

поэтому здесь критическая поверхность одномерна. Если $\bar{g}(\mu)$ находится на этой поверхности, причем $\xi = \xi_0 > 0$ при $\mu = \mu_0$, то при больших значениях μ для $\bar{g}$ справедлива формула (15), в которой

$$\xi = \xi_0 \left[1 + \xi_0 \sum_i a_i^2 \ln(\mu/\mu_0)\right]^{-1}.$$

Мы видим, что в этом случае $\bar{g}_i$ плавно подходит к g_i^* при $\mu \to \infty$, и теория является асимптотически безопасной. С другой стороны, если $\bar{g}_i$ не лежит на линии (15), то возникает расходимость при *конечном* значении μ. Для $\bar{g}_i \to \infty$ решением уравнения (14) будет

$$\bar{g}_i \to a_i \left[-\sum a_j^2 \ln(\mu/\mu_\infty)\right]^{-1},$$

и мы видим, что $\bar{g}_i \to \infty$, когда μ приближается к μ_∞. Если предположить, что бесконечность в параметрах связи будет давать нефизическую особенность в скоростях реакций, то мы должны здесь заключить, что теория, связи которой не лежат на ультрафиолетовой критической поверхности (15), будет обнаруживать нефизические особенности при энергии $E = \mu_\infty$.

Конечно, вопрос о том, предвещает ли бесконечность в константах связи особенность в скоростях реакций, зависит от того, как определяются константы связи. Мы всегда могли бы принять необычное определение (например, $\bar{g}' \equiv (\bar{g} - g^*)^{-1}$), такое, чтобы скорости реакций были конечны даже при бесконечных параметрах связи. Можно избежать этой проблемы, если мы определим кон-

---

[1]) Это является обобщением примера, предложенного мне Виттеном. Конечно, пример специально подобран так, чтобы иметь траектории с желаемым поведением, т. е. они попадают в фиксированную точку для $\mu \to \infty$ или уходят на бесконечность для конечных значений μ. Однако именно этот пример показывает, что такое поведение может иметь место для β-функций, которые вообще не являются патологическими.

станты связи как коэффициенты разложения в степенной ряд самих скоростей реакций около некоторой физической точки перенормировки. В наинизшем порядке теории возмущений эта процедура неотличима от обычной процедуры перенормировки, в которой $\bar{g}_i(\mu)$ определяются в терминах разложения в степенные ряды функций Грина около некоторой точки перенормировки вне массовой поверхности. По-видимому, имеет смысл исследовать, можно ли было бы получить последовательное определение перенормированных параметров связи в терминах скоростей реакций, а не функций Грина, но здесь мы не будем пытаться этого сделать.

Число свободных параметров в асимптотически безопасной теории равно просто размерности ультрафиолетовой критической поверхности. Если критическая поверхность бесконечномерна, то требование, чтобы физическая теория лежала на этой поверхности, оставляет нам бесконечное число неопределенных параметров, и мы мало чего достигаем. В другом пределе, если размерность критической поверхности нулевая, требование асимптотической безопасности не может быть удовлетворено вообще. Предположим, что размерность критической поверхности — некоторое конечное число C; в этом случае теория будет иметь C свободных параметров, из них $C-1$ являются безразмерными параметрами, которые определяют отдельную траекторию на C-мерной критической поверхности, а один параметр является размерным и указывает нам значение μ, при котором достигается некоторая заданная точка на этой траектории. Лучшим вариантом, конечно, был бы случай $C=1$; при этом физика не имела бы никаких свободных параметров, кроме одной размерной постоянной, которая просто определяла бы наши единицы массы или длины. Пока величина C конечна, условие асимптотической безопасности будет для нас играть такую же роль, как условие перенормируемости в квантовой электродинамике: оно служит для того, чтобы фиксировать все, кроме, может быть, конечного числа параметров теории. Фактически, как мы увидим, условие асимптотической безопасности будет в некоторых случаях требовать, чтобы теория была перенормируемой в обычном смысле.

Размерность критической поверхности можно определить, исходя из поведения функций $\beta_i(\bar{g})$ вблизи фиксированной точки. В окрестности g^* (12) можно записать в виде

$$\mu \frac{d}{d\mu} \bar{g}_i(\mu) \to \sum_j B_{ij} [\bar{g}_j(\mu) - g_j^*], \tag{16}$$

где

$$B_{ij} \equiv (\partial\beta_i(\bar{g})/\partial\bar{g}_j)_{\bar{g}=g^*}. \tag{17}$$

Общим решением является

$$\bar{g}_i(\mu) \to \sum_K C_K V_i^K \mu^{\lambda_K} + g_i^*, \tag{18}$$

где V^K — собственный вектор B_{ij} с собственным значением λ_K,

$$\sum_j B_{ij} V_j^K = \lambda_K V_i^K, \qquad (19)$$

а C_K — произвольные коэффициенты. Очевидно, условием того, чтобы $\bar{g}_i(\mu)$ достигало g_i^* при $\mu \to \infty$, является условие обращения C_K в нуль для всех возможных собственных значений $\lambda_K > 0$. (Возможность нулевых собственных значений здесь является неудобством, к которому мы вернемся позднее.) Размерность ультрафиолетовой критической поверхности тогда будет равна числу остающихся параметров C_K, т. е. числу отрицательных собственных значений B_{ij}.

В таком случае решающая проблема состоит в том, чтобы определить, сколько собственных значений B-матрицы являются отрицательными. Во всех случаях, которые мне известны, это число конечно. Это можно пояснить наводящими, хотя и весьма нестрогими соображениями. Вспомним, что

$$\bar{g}_i(\mu) \equiv \mu^{-d_i} g_i(\mu),$$

где d_i — размерность зависимости от массы перенормированной константы связи $g_i(\mu)$, не подвергшейся изменению масштаба. Зависимость $g_i(\mu)$ от μ возникает из зависимости петлевых графиков от импульсов в точке перенормировки, так что

$$\beta_i = -d_i \bar{g}_i + \text{петлевые вклады} \qquad (20)$$

и

$$B_{ij} = -d_i \delta_{ij} + \text{петлевые вклады}. \qquad (21)$$

Тогда добавление к взаимодействию производных и степеней полей будет всегда *понижать* размерность d_i, поэтому не более чем конечное число d_i может быть положительным и все, кроме конечного числа, будут лежать ниже любого заданного отрицательного значения. В отсутствие «петлевых вкладов» собственные значения B_{ij} были бы как раз величинами $-d_i$, из которых все, кроме конечного числа, положительны. «Петлевые вклады» могут, конечно, изменить знаки некоторых из собственных значений B_{ij}, но если эти вклады являются связанными, то они не могут изменить знак бесконечного числа больших положительных собственных значений, и только конечное число может быть отрицательно. Таким образом, мы можем предположить, что ультрафиолетовая критическая поверхность будет вообще иметь конечную размерность.

Этот вывод в ряде случаев эмпирически подтверждается наблюдаемым существованием фазовых переходов второго рода. Фазовый переход второго рода будет, вообще говоря, происходить при значениях параметров теории, при которых стремятся к нулю массы или расходятся корреляционные длины, так что физические величины могут обнаруживать масштабно-инвариантное поведение при

больших расстояниях или малых импульсах. Повторяя соображения, приведенные в этом разделе, мы видим, что такой скейлинг должен быть связан с фиксированной точкой g^*, где $\beta_i(g^*)$ стремится к нулю; фазовый переход происходит, когда параметры теории принимают значения на *инфракрасной* критической поверхности [1]), состоящей из траекторий (12), которые попадают в фиксированную точку при $\mu \to 0$. Из (18) мы видим, что число параметров, которые следует установить, чтобы использовать теорию на инфракрасной критической поверхности, равно числу отрицательных собственных значений B_{ij} и, следовательно, равно именно размерности *ультрафиолетовой* критической поверхности. Но во всех случаях мы знаем, что это число конечно: только конечное число параметров (температура, давление, магнитное поле) должны быть установлены, чтобы ввести фазовый переход второго рода. *Таким образом, по крайней мере для фиксированных точек, связанных с фазовыми переходами второго рода, мы можем быть уверены, что ультрафиолетовая критическая поверхность имеет конечную размерность.*

Мы уже отмечали, что для конечномерной критической поверхности условие асимптотической безопасности действует весьма аналогично условию перенормируемости, ограничивая свободные параметры в физических теориях. Фактически теперь мы можем видеть, что связь между асимптотической безопасностью и перенормируемостью является даже более тесной. Любая теория всегда будет иметь фиксированную точку в начале, $g^*=0$. (Если существенные связи стремятся к нулю при одном масштабе перенормировки μ, они будут стремиться к нулю при всех μ, поэтому $\beta_i(\bar{g})$ всегда обращается в нуль при $\bar{g}=0$.) Предположим, что для некоторой теории она является единственной подходящей фиксированной точкой с ультрафиолетовой критической поверхностью ненулевой размерности, так что асимптотическая безопасность требует, чтобы связи лежали на этой поверхности. «Петлевой вклад» в (21) обращается в нуль при $\bar{g}_i=0$, поэтому B-матрица для этой фиксированной точки есть

$$B_{ij} = -d_i\delta_{ij}. \tag{22}$$

Таким образом, чтобы траектория $\bar{g}_i(\mu)$ попадала в точку $g^*=0$ при $\mu \to \infty$, необходимо, чтобы все g_i при $d_i<0$ обращались в нуль. Но именно эти взаимодействия являются неперенормируемыми [2]), поэтому такой тип теории должен быть перенормируемым в обычном смысле.

Строго говоря, перенормируемость может быть недостаточна. B-матрица для $g^*=0$ будет, вообще говоря, иметь несколько нулевых собственных значений, соответствующих «строго» перенормируемым взаимодействиям с $d_i=0$, и необходимо также, чтобы эти вза-

---

[1]) См. примечание на с. 410.
[2]) См. примечание 2 на с. 407.

имодействия имели $\beta_i/g_i<0$ вблизи $\bar{g}_i=0$, чтобы $\bar{g}_i(\mu)$ стремились к нулю при $\mu\to\infty$. Следовательно, ультрафиолетовая критическая поверхность фиксированной точки $g^*=0$ фактически состоит из всех теорий, которые перенормируемы *и* асимптотически свободны[1]). Однако с практической точки зрения перенормируемая теория типа квантовой электродинамики может считаться асимптотически безопасной, даже если она не является асимптотически свободной, поскольку рост строго перенормируемой связи типа $e(\mu)$ является только логарифмическим, и какие-либо нефизические особенности могли бы появляться только при экспоненциально высоких энергиях. Это было бы не так, если бы теория включала неперенормируемые взаимодействия типа взаимодействия Ферми.

В некоторых случаях асимптотическая безопасность может вести к перенормируемости, даже если фиксированная точка имеется не при $g^*=0$. Известная работа [2]) Гелл-Манна и Лоу говорит о существовании фиксированной точки $e^*\neq0$ в квантовой электродинамике, где $\beta_e(e^*)=0$. Если существует такая фиксированная точка в квантовой электродинамике, то она является фиксированной точкой для самой общей полевой теории фотонов и электронов, но нет никаких оснований ожидать, что траектории с не обращающимися в нуль значениями для неперенормируемых связей должны попадать в эту фиксированную точку. Поэтому, исключая другие фиксированные точки, действие условия асимптотической безопасности будет требовать именно перенормируемости обычного типа.

Проблема, перед которой ставит нас рассмотрение квантовой гравитации, состоит в том, что может не быть никаких теорий, которые являются перенормируемыми, оставаясь только асимптотически свободными. Поэтому мы должны искать другие фиксированные точки вдали от $g^*=0$. Вообще говоря, не имеется особых причин, по которым фиксированная точка с $g^*\neq0$ имела бы малое g^*, поэтому теория возмущений не может быть нам слишком полезной для поиска таких фиксированных точек или исследования их свойств.

Почти такая же проблема возникла в теории критических явлений: несостоятельность обычной теории поля показала, что фазовые переходы не связаны с фиксированной точкой при нулевой связи, и возникла необходимость искать другие фиксированные точки. В этом случае проблема могла бы быть решена [130, 131] продолжением по пространственной размерности системы, «ε-расширением». Поэтому представляется разумным попытаться рассмотреть продолжение по пространственно-временнóй размерности в исследуемой проблеме.

Предположим, что мы могли бы найти некоторую размерность пространства-времени $D_r<4$, при которой существует перенорми-

[1]) Асимптотическая свобода неабелевых калибровочных теорий была открыта Политцером [128], а также Гроссом и Вильчеком [129].

[2]) См. примечание на с. 410.

руемая и асимптотически свободная теория гравитации. Как мы видели, это означало бы, что фиксированная точка при $g^*=0$ имеет ультрафиолетовую критическую поверхность конечной размерности. Если мы затем увеличим размерность D, теория будет становиться неперенормируемой, но непрерывность заставляет нас ожидать, что фиксированная точка, которая имеется при $g^*=0$ для $D=D_r$, будет плавно смещаться и что по крайней мере для конечной области значений D выше D_r ультрафиолетовая критическая поверхность будет сохранять ту же размерность. Следовательно, для $D=D_r+\varepsilon$ мы можем рассчитывать найти фиксированную точку с g^* порядка ε и изучить свойства этой фиксированной точки с помощью разложения по степеням ε.

Этот подход уже применялся [132—135] к нелинейной σ-модели (которая является перенормируемой и асимптотически свободной при $D_r=2$), хотя с несколько иной точки зрения. Его применение к гравитации будет описано в разд. 6 с использованием методов, обсуждаемых в разд. 5.

При проведении вычислений очень полезно знать, что хотя β_i-функции и B_{ij}-матрица зависят от деталей нашей процедуры перенормировки и от того, как определяются параметры связи g_i, однако *собственные значения* B_{ij}-матрицы от этого не зависят [1]. Имеется широкий выбор способов, которыми мы могли бы изменить определение $g_i(\mu)$.

а. Мы могли бы просто выбрать другой набор точек перенормировки, и в этом случае переопределенные масштабно параметры связи (9) стали бы функциями $\tilde{g}_i(\tilde{\mu})$ масштаба импульсов $\tilde{\mu}$ в новых точках перенормировки.

б. Мы можем использовать размерную регуляризацию [137 — 139], и в этом случае «перенормированные» связи могут быть приняты за постоянные члены в разложении неперенормированных связей g_{i0} в ряд Лорана вблизи пространственно-временно́й размерности $D=4$:

$$g_{i0} \xrightarrow[D\to 4]{} \tilde{\mu}^{d_i(D)}\left[\tilde{g}_i(\tilde{\mu})+\sum_{\nu=1}^{\infty}\frac{b_{\nu i}(\tilde{g}(\tilde{\mu}))}{(D-4)^{\nu}}\right], \tag{23}$$

где $\tilde{\mu}$ — «единица массы», вводимая для того, чтобы сделать $\tilde{g}_i(\tilde{\mu})$ безразмерными.

в. Мы можем ввести обычное ультрафиолетовое обрезание при импульсе $\tilde{\mu}$ и взять функции $\tilde{g}_i(\tilde{\mu})$ в качестве переопределенных непе-

---

[1]) Это показано для теории критических явлений Вегнером [136]. В этой работе подчеркивается, что инвариантность критических экспонент относится только к тем собственным векторам, которые связаны с существенными связями. В статистической физике связи, которые не существенны, называются «излишними». Я использую здесь вместо этого термин «несущественный», поскольку, обращаясь к противоположному ему термину, легче сказать «существенный», чем «не излишний».

ренормированных связей, причем выбрать зависимость обрезания таким образом, чтобы скорости реакций не зависели от обрезания [1]).

В этих или других случаях новые связи $\tilde{g}_i$ должны быть представимы как функции старых связей $\tilde{g}_j(\mu)$ и единственной другой безразмерной величины $\tilde{\mu}/\mu$:

$$\tilde{g}_i(\tilde{\mu}) = \tilde{g}_i\left(\frac{\tilde{\mu}}{\mu},\ \bar{g}_i(\mu)\right).$$

Однако новые связи не зависят от того, как определены старые связи, поэтому они являются не зависящими от μ:

$$0 = \mu\frac{\partial\tilde{g}_i}{\partial\mu} = -\tilde{\mu}\frac{\partial\tilde{g}_i}{\partial\tilde{\mu}} + \sum_j \frac{\partial\tilde{g}_i}{\partial\bar{g}_j}\mu\frac{d\bar{g}_i}{d\mu}.$$

Другими словами, мы можем определить новую β_i-функцию

$$\tilde{\beta}_i(\tilde{g}(\tilde{\mu})) \equiv \tilde{\mu}\frac{\partial\tilde{g}_i}{\partial\tilde{\mu}}, \tag{24}$$

которая связана со старой β-функцией законом преобразования контравариантного вектора

$$\tilde{\beta}_i(\tilde{g}) = \sum_j \frac{\partial\tilde{g}_j}{\partial\bar{g}_j}\beta_j(\bar{g}). \tag{25}$$

Мы видим, что существование фиксированной точки является инвариантным понятием: если $\beta_i(g^*)=0$, то $\tilde{\beta}_i(\tilde{g}(g^*))=0$. Сами β-функции не являются инвариантными, и так же не инвариантны их производные

$$\frac{\partial\tilde{\beta}_i}{\partial\tilde{g}_j} = \sum_{kl}\frac{\partial^2\tilde{g}_i}{\partial\bar{g}_k\,\partial\bar{g}_l}\frac{\partial\bar{g}_l}{\partial\tilde{g}_j}\beta_k(\bar{g}) + \sum_{kl}\frac{\partial\tilde{g}_i}{\partial\bar{g}_k}\frac{\partial\beta_k(\bar{g})}{\partial\bar{g}_l}\frac{\partial\bar{g}_l}{\partial\tilde{g}_j}.$$

Но в фиксированной точке первый член обращается в нуль, поэтому новая B-матрица связана со старой следующим образом:

$$\tilde{B}_{ij} = \sum_{kl} A_{ik}B_{kl}A^{-1}_{lj}. \tag{26}$$

$$A_{ik} \equiv \left(\frac{\partial\tilde{g}_i(\bar{g})}{\partial\bar{g}_k}\right)_{\bar{g}=g^*}. \tag{27}$$

Это преобразование подобия, поэтому собственные значения $\tilde{B}$ — те же самые, что и собственные значения B. (Эти собственные значения известны как *критические экспоненты*; они зависят только от типа степеней свободы системы, а не от каких-либо других физических переменных.) В частности, вопрос об асимптотической безо-

[1]) Это метод, используемый в большинстве приложений ренорм-группы к критическим явлениям. См., например, работу [23].

пасности является вопросом, который может быть адресован к любому из формализмов «а», «б», «в», охарактеризованных выше, с уверенностью, что ответ будет тем же самым.

4. ФИЗИКА ПРИ ОБЫЧНЫХ ЭНЕРГИЯХ

Условие асимптотической безопасности требует существования фундаментального масштаба энергии M: это значение μ, при котором траектория $\bar{g}_i(\mu)$ приближается в пределах некоторого определенного расстояния к фиксированной точке g_i^*. Ниже мы увидим, что эта характерная энергия будет порядка планковской массы [1]), $M_p=1{,}2\cdot 10^{19}$ ГэВ. Проблема, к которой мы обращаемся теперь, состоит в описании физических явлений при обычных энергиях, которые гораздо меньше чем M. В частности, мы хотим узнать, почему гравитационные взаимодействия на макроскопических расстояниях так хорошо описываются эйнштейновским лагранжианом — $\sqrt{g}R/16\pi G$ и почему слабые, сильные и электромагнитные взаимодействия должны при обычных энергиях так хорошо описываться перенормируемыми квантовыми полевыми теориями [2]).

Рассмотрим сначала случай чистой гравитации с лагранжианом вида (8). Предположим, что теория является асимптотически безопасной, поэтому бесконечное множество связей должно обязательно лежать на траектории $\bar{g}_i(\mu)$, которая попадает в некоторую фиксированную точку g_i^* при $\mu\to\infty$. Из ренорм-групповых соображений предыдущего раздела следует, что физика при низких энергиях определяется поведением связей $\bar{g}_i(\mu)$ при $\mu\to 0$. Простейшая возможность для $\bar{g}_i(\mu)$ состоит в приближении в этом пределе к другой фиксированной точке. Далее, в предыдущем разделе мы видели, что в дополнение к g_i^* всегда существует фиксированная точка в нуле. К тому же каждый член в лагранжиане (8) является неперенормируемым, поэтому каждое собственное значение матрицы $\partial\beta_i/\partial\bar{g}_j$ при $\bar{g}_i=0$ является положительным. Это означает, что фиксированная точка в нуле является полностью отталкивающей для $\mu\to\infty$, но она, кроме того, является полностью притягивающей для $\mu\to 0$: в каком бы направлении мы ни двигались от нуля, мы остаемся на инфракрасной критической поверхности. Отсюда следует, что имеется по крайней мере конечная область вокруг нуля, в пределах которой *любая* траектория притягивается к началу при $\mu\to 0$. Мы не знаем, лежит ли ультрафиолетовая фиксированная точка g_i^* в пределах этой области, но нет ничего неприемлемого в предположении, что лежит. При этом предположении мы можем заключить, что все $\bar{g}_i(\mu)$ обращаются в нуль для $\mu\to 0$.

---

[1]) См. примечание 1 на с. 409.

[2]) Обзоры современной перенормируемой калибровочной теории слабых и электромагнитных взаимодействий см. в работах [140, 141].

(Здесь мы неявно предположили, что лагранжиан не содержит члена с космологической постоянной $\Lambda\sqrt{g}$. Такой член был бы сверхперенормируемым и поэтому соответствовал бы инфракрасному отталкивательному собственному вектору $\partial\beta_i/\partial\bar{g}_j$ при $\bar{g}_i=0$. Космологическая постоянная не требуется в качестве контрчлена в чистой гравитации, по крайней мере в схеме размерной регуляризации. Однако теории с массивными частицами будут, вообще говоря, требовать контрчлена с космологической постоянной, и есть что-то загадочное в том, почему Λ, например, не больше на сорок порядков, чем наблюдаемый верхний предел.)

Для $\bar{g}_i(\mu)$, близких к нулю, «петлевой вклад» в (20) пренебрежимо мал, и решение (12) становится просто

$$\bar{g}_i(\mu) \approx (M_i/\mu)^{d_i} \quad \text{для} \quad \mu \ll M_i,$$

где M_i— набор неизвестных постоянных интегрирования. Величины M_i связаны друг с другом условием асимптотической безопасности, так что значение $\bar{g}(\mu)$ должно лежать на ультрафиолетовой критической поверхности фиксированной точки g^*. Для одномерной критической поверхности условие асимптотической безопасности оставляет только один свободный параметр — фундаментальный энергетический масштаб M, поэтому все M_i должны быть порядка M. То же самое верно, даже если размерность критической поверхности C больше единицы, при условии что $(C—1)$ параметров, определяющих траекторию $\bar{g}_i(\mu)$, не принимают исключительно больших или исключительно малых значений. Тогда мы приходим к выводу, что первоначальные связи до изменения масштаба (9) имеют порядок величины

$$g_i(\mu) \approx M^{d_i} \quad \text{для} \quad \mu \ll M. \tag{28}$$

В частности, (28) дает гравитационную константу G порядка M^{-2}, поэтому M должно быть порядка планковской массы

$$M \approx M_P \equiv G^{-1/2} = 1{,}2\cdot 10^{19}\ \text{ГэВ}. \tag{29}$$

Теперь мы можем понять, почему гравитационные явления так хорошо описываются теорией Эйнштейна на макроскопических расстояниях. Рассмотрим связанную функцию Грина для системы гравитационных полей в точках с типичными пространственно-временными расстояниями r. Ультрафиолетовые расходимости устраняются перенормировкой бесконечного множества параметров связи, однако здесь существенно, что перенормированные параметры связи $g_i(\mu)$ будут определяться в точках перенормировки с импульсами порядка $1/r$, а не M_P. При этом подходе интегралы в диаграммах Фейнмана будут начинать сходиться при импульсах порядка $1/r$, поэтому r будет единственным размерным параметром в теории, отличным от самих параметров связи. Уравнения (28) и (29) пока-

зывают, что константы связи в диаграмме с $\mathcal{N}_i$ вершинами типа i дают множитель, пропорциональный N-й степени $G^{1/2}$ или M_P^{-1}, где

$$N = -\sum_i \mathcal{N}_i d_i, \tag{30}$$

поэтому из обычного размерного анализа ясно, что вклад такой диаграммы будет подавляться множителем

$$(G^{1/2}/r)^N = (rM_P)^{-N}. \tag{31}$$

Таким образом, основными диаграммами для значений r, гораздо больших, чем планковская длина $M_P^{-1} = 1{,}6 \cdot 10^{-33}$ см, будут диаграммы с наименьшими значениями N. Размерность d_i определяется как

$$d_i = 4 - p_i - g_i, \tag{32}$$

где p_i — число производных и g_i — число гравитонных полей во взаимодействии типа i. Чтобы вычислить N, используем известные топологические соотношения

$$\mathcal{L} = \mathcal{l} - \sum_i \mathcal{N}_i + 1, \tag{33}$$

$$2\mathcal{l} + \mathcal{E} = \sum_i \mathcal{N}_i g_i, \tag{34}$$

где $\mathcal{L}$ — число петель; $\mathcal{l}$ — число внутренних линий; $\mathcal{E}$ — число внешних линий в диаграмме. Вместе с (30) и (32) они дают

$$N = \sum_i \mathcal{N}_i (p_i - 2) + 2\mathcal{L} + \mathcal{E} - 2. \tag{35}$$

Если не рассматривать никакого сверхперенормируемого космологического члена $\Lambda\sqrt{g}$, то гравитационными взаимодействиями с наименьшим числом производных являются взаимодействия, выводимые из эйнштейновского члена $-\sqrt{g}R/16\pi G$, для которых $p_i = 2$. Следовательно, для любой заданной функции Грина с заданным числом внешних линий $\mathcal{E}$ основными диаграммами для $r \gg M_P^{-1}$ будут древесные диаграммы ($\mathcal{L} = 0$), построенные только из эйнштейновского лагранжиана. Суммирование этих древесных диаграмм равносильно решению классических полевых уравнений [1]); в частности, одногравитонная функция Грина в присутствии классического фонового распределения энергии и импульса удовлетворяет классическим полевым уравнениям Эйнштейна для гравитационного поля, создаваемого этим тензором энергии-импульса. Следующие поправки будут появляться как от диаграмм с одной петлей в чистой общей теории относительности, так и в равной степени от древесных диаграмм, содержащих *одну* вершину, возникающих из более высоких

[1]) Это показано для обычных полевых теорий в работах [142—144] и специально для общей теории относительности в низших порядках в работе [145]. Приложения этих результатов к движению больших масс рассмотрены в работе [146].

взаимодействий $\sqrt{g}R^2$ или $\sqrt{g}R^{\mu\nu}R_{\mu\nu}$, для которых $p_i=4$. Ультрафиолетовая расходимость в петлевом графике исключается контрчленом, который дают взаимодействия $\sqrt{g}R^2$ и $\sqrt{g}R^{\mu\nu}R_{\mu\nu}$. Из уравнения (35) видно, что эти квантовые поправки к классической общей теории относительности подавляются множителем порядка $(rM_P)^{-2}$ и, следовательно, полностью пренебрежимы на макроскопических расстояниях.

Кстати, из (35) видно, что даже вклады классических древесных диаграмм в заданную функцию Грина подавляются множителем, пропорциональным $G^{1/2}$ для каждой добавочной внешней гравитонной линии. Когда мы вычисляем метрику, создаваемую массой m, мы также получаем другой множитель $G^{1/2}m$ для каждой связи этих внешних линий с массой. Причина, по которой обмен *древесными диаграммами* гравитонных линий при $\mathscr{E}>2$ внешних линий оказывает обнаружимое влияние на движение планет, заключается как раз в том, что солнечная масса $m_\odot$ так велика, что $Gm_\odot/r$ уже не является полностью пренебрежимо малой величиной.

Приведенное выше обсуждение может быть непосредственно распространено на теории гравитации и материи по крайней мере в случае, когда не имеется никаких масс и никаких сверхперенормируемых или асимптотически свободных перенормируемых взаимодействий между материальными полями. (Например, она была бы уместна для теории гравитонов, фотонов и безмассовых электронов.) В такой теории фиксирования точка при нулевой связи является еще полностью притягивающей для $\mu\to 0$, поэтому, согласно нашим предыдущим аргументам, связи при обычной энергии будут иметь порядок величины

$$g_i(\mu)=O\left(M_P^{d_i}\right) \quad \text{для } \mu\ll M_P.$$

Физические явления при обычных энергиях $E\ll M_P$ будут полностью управляться перенормируемыми взаимодействиями, для которых $d_i=0$, поскольку все неперенормируемые взаимодействия с $d_i<0$ подавляются как степени E/M_P и, следовательно, полностью необнаружимы. Одна лишь гравитация составляет исключение: как мы видели, тот факт, что гравитация когерентно связана с каждой частицей в большом теле, подобном Солнцу, позволяет нам наблюдать ее макроскопические проявления, несмотря на ее собственную слабость.

Теории гравитации и материи, включающие массы или асимптотически свободные перенормируемые взаимодействия, требуют несколько большего внимания. Во-первых, могут быть частицы с массой порядка M_P — например, имеются промежуточные векторные бозоны (почти столь же тяжелые) в некоторых суперобъединенных теориях слабых, электромагнитных и сильных взаимодействий [18]. Такие частицы не вызывают здесь никаких проблем; в функциях Грина при обычной энергии $E\ll M_P$ любая внутренняя линия с массой M порядка M_P может быть заменена рядом неперенормиру-

емых взаимодействий, полученным разложением пропагаторов тяжелых частиц:

$$\frac{1}{q^2+M^2}=\frac{1}{M^2}-\frac{q^2}{M^4}+\frac{(q^2)^2}{M^6}-\dots .$$

(Здесь опять существенно, чтобы масштабы импульсов μ всех точек перенормировки были приняты величинами порядка E, а не M, так чтобы интегралы сходились при импульсах $q\approx E\ll M$.) Таким способом мы получаем эффективную полевую теорию [1]), включающую только легкие частицы с массами, гораздо меньшими чем M_p. Эти частицы, конечно, содержат в себе все знакомые нам: гравитоны, фотоны, лептоны, кварки, промежуточные векторные бозоны, хиггсовские бозоны и глюоны [148, 149].

Теперь предположим, что эта эффективная полевая теория включает либо некоторые малые массы $m\ll M_p$, либо некоторые асимптотически свободные, перенормируемые взаимодействия, действующие между легкими частицами, либо и то и другое. В этом случае фиксированная точка в начале в пространстве константы связи не является больше полностью притягивающей для $\mu\to 0$. Однако мы можем теперь определить инфракрасно-притягивающую *область*, или даже инфракрасно-притягивающую *точку*. Обозначим масштабно переопределенные перенормируемые связи (включая массы) через $\bar{f}_a(\mu)$, а неперенормируемые связи — через $\bar{F}_A(\mu)$. Поскольку перенормируемые связи никогда не порождают бесконечностей, которые требуют неперенормируемых контрчленов, уравнения ренорм-группы для $\bar{F}_A(\mu)\ll 1$ должны принимать вид

$$\mu\frac{d}{d\mu}\bar{f}_a(\mu)\approx\beta_a(\bar{f}(\mu)),$$

$$\mu\frac{d}{d\mu}\bar{F}_A(\mu)\approx\sum_B B_{AB}(\bar{f}(\mu))\,\bar{F}_B(\mu).$$

Кроме того, для $\bar{f}(\mu)\ll 1$ петлевые вклады в B_{AB} становятся пренебрежимыми, так что

$$B_{AB}(\bar{f})\approx -d_A\delta_{AB}\quad(\bar{f}\ll 1).$$

По определению F_A является неперенормируемым взаимодействием, поэтому $d_A>0$. Таким образом, имеется по крайней мере конечная область $\mathcal{A}$ на поверхности $\bar{F}_A=0$, в пределах которой все собственные значения матрицы B_{AB} положительны. Это инфракрасно-притягивающая область (всякая траектория на по крайней мере конечном участке плоскости $\mathcal{R}_{\mathcal{A}}$ вокруг $\mathcal{A}$ притягивается к ней при $\mu\to 0$). Кроме того, мы не знаем, лежит ли ультрафиолетовая фиксированная точка g_i^* внутри области $\mathcal{R}_{\mathcal{A}}$, однако не столь уж неразумно пред-

[1]) Это — обобщение результата Апплеквиста и Караццоне [147]. Они показали, что в перенормируемой теории «легких» и (гораздо более) «тяжелых» частиц взаимодействия между «легкими» частицами даются эффективной перенормируемой теорией поля, в которой «тяжелые» частицы не появляются.

положить, что лежит. На самом деле, это, по-видимому, может быть проверено экспериментально: имеется огромная область масштабов перенормировки 1 ГэВ$\ll\mu\ll10^{19}$ ГэВ, внутри которой масштабно переопределенные неперенормируемые гравитационные связи $\bar{F}_A(\mu)$ чрезвычайно малы, а перенормируемые слабые, электромагнитные и «сильные» взаимодействия также имеют довольно малые связи, так что B_{AB} имеет положительные собственные значения $\approx -d_A$. Если предположить, что ультрафиолетовая фиксированная точка g_i^* действительно лежит внутри области $\mathcal{R}_{\mathcal{A}}$, то мы можем снова заключить, что влияние неперенормируемых связей становится пренебрежимым при обычных энергиях. Более конкретно, в пределах области масштабов перенормировки, в которой все $\bar{g}_i(\mu)$ малы, непереопределенные масштабно связи $g_i(\mu)$ опять будут порядка M^{d_i}, как в (28), где M — обычная постоянная интегрирования, связанная с масштабом перенормировки, при котором $\bar{g}_i(\mu)$ приближается к фиксированной точке g_i^*. Обычные значения гравитационных связей G и другие неперенормируемые связи определяются в точке перенормировки с μ порядка массы типичной легкой частицы, где масштабно переопределенные связи (включая m/μ) становятся как раз порядка единицы, поэтому эти обычные связи также будут порядка M^{d_i}. Мы можем, таким образом, отождествить M с планковской массой, как в (29), и снова сделать вывод, что влияние неперенормируемых связей при обычных энергиях $E\ll M_P$ будет подавляться, как степени E/M_P.

5. ПРОДОЛЖЕНИЕ ПО РАЗМЕРНОСТИ

В разд. 3 было подчеркнуто, что существование фиксированной точки и размерность ее ультрафиолетовой критической поверхности не зависят от того, определяем ли мы параметры связи обычной перенормировкой, размерной регуляризацией или «плавающим» ультрафиолетовым обрезанием. Однако опыт показывает, что метод размерной регуляризации [137—139] с большим запасом является самым удобным для актуальных вычислений [1]). Несколько неожиданным оказывается, что размерная регуляризация дает также очень удобный базис для изучения фиксированных точек при произвольной, нецелой размерности [2]); в разд. 3 мы отмечали, что такое

[1]) Использование размерной регуляризации для построения новой системы уравнений ренорм-группы принадлежит т'Хофту [150]. Вывод, приведенный здесь, представляет собой несколько упрощенную версию вывода, данного т'Хофтом.

[2]) В работе [150] перенормированная константа связи была определена прямо при $D=4$ как постоянный член в разложении в ряд Лорана в окрестности точки $D=4$. Расширение этого определения и соответствующего уравнения ренорм-группы на произвольную иррациональную пространственно-временную размерность неявно использовалось в течение некоторого времени в работах [23, 153, 154] и явно было дано Гроссом [151].

продолжение по размерности дает метод, при котором можно использовать теорию возмущений для изучения фиксированных точек.

Размерная регуляризация позволяет нам вычислить все фейнмановские интегралы в конечном виде для нерациональной пространственно-временно́й размерности D, но эти интегралы будут иметь полюсы при приближении D к различным рациональным значениям. Возьмем некоторый определенный набор рациональных значений пространственно-временно́й размерности D_s и предположим, что неперенормированные константы связи $g_{i0}(D)$ имеют полюсы, которые будут аннулировать полюсы при $D=D_s$ в фейнмановских интегралах, давая скорости реакций, которые конечны при $D=D_s$. (В оригинальной формулировке [150] этого подхода набор D_s состоял из единственной физической размерности пространства-времени $D=4$; обобщение на несколько D_s представлено здесь потому, что оно не требует в сущности никакой дополнительной работы и могло бы оказаться полезным.) Чтобы выразить $g_{i0}(D)$ в терминах перенормированных параметров связи, которые являются безразмерными для всех D, мы должны ввести единицу массы μ и записать разложение в ряд Лорана для масштабно-переопределенной связи $g_{i0}(D)\mu^{-d_i(D)}$ вместо обычной неперенормированной связи $g_{i0}(D)$. (Как обычно, $d_i(D)$ — размерность $g_{i0}(D)$ в степени массы при размерности пространства-времени D.) Кроме полюсов при $D=D_s$ это разложение будет иметь остаточный член, который аналитичен по D и который мы просто *определяем* как безразмерную связь $g_i(\mu, D)$. Коэффициенты $b_{\nu i}^{(s)}$ полюсов порядка ν в $g_{i0}(D)\mu^{-d_i(D)}$ при $D=D_s$ будут зависеть от μ и D только через их зависимость от перенормированной связи $g_i(\mu, D)$, поскольку не имеется никакого размерного параметра, с которым можно было бы сравнить μ, и поскольку любая отдельная зависимость $b_{\nu i}^{(s)}$ от D меняла бы только аналитические члены и полюсы более низкого порядка по D. Поэтому разложение в ряд Лорана может быть записано в виде

$$g_{i0}(D)\mu^{-d_i(D)}=g_i(\mu, D)+\sum_s\sum_{\nu=1}^{\infty}(D-D_s)^{-\nu}b_{\nu i}^{(s)}(g(\mu, D)). \quad (36)$$

(Теперь мы опускаем тильду, использованную в разд. 3, чтобы отличить $g_i(\mu, D)$ от обычной масштабно-переопределенной перенормируемой связи $\bar{g}_i(\mu)$.)

Чтобы вычислить функции Гелл-Манна — Лоу $\beta_i(g, D)$, сначала дифференцируем по μ и находим

$$-d_i(D)\left[g_i+\sum_{s,\nu}(D-D_s)^{-\nu}b_{\nu i}^{(s)}(g)\right]=$$
$$=\beta_i(g, D)+\sum_{s,\nu,j}(D-D_s)^{-\nu}b_{\nu ij}^{(s)}(g)\beta_j(g, D), \quad (37)$$

где

$$\beta_i(g(\mu, D), D)\equiv\mu\frac{\partial}{\partial\mu}g_i(\mu, D) \quad (38)$$

и

$$b_{\nu ij}^{(s)}(g) \equiv \partial b_{\nu i}^{(s)}(g)/\partial g_j. \tag{39}$$

Кроме того, мы запишем β_i как функцию от всех $g_j(\mu, D)$, а также от D, но не от μ, поскольку здесь нет никакой размерной величины, с которой можно было бы сравнить μ.

Тогда размерность $d_i(D)$ величины g_{i0} всегда является линейной функцией пространственно-временно́й размерности D, которую мы будем записывать как [150]

$$d_i(D) = \sigma_i + \rho_i D. \tag{40}$$

Левая часть (37) может быть теперь переписана в виде

$$-\rho_i g_i D - \Big[\sigma_i g_i + \sum_s b_{1i}^{(s)}(g)\rho_i\Big] - \\ - \sum_{s,\,\nu}(D-D_s)^{-\nu}\,[\rho_i b_{\nu+1,i}^{(s)}(g) + (\sigma_i + \rho_i D_s)\, b_{\nu,i}^{(s)}(g)], \tag{41}$$

где суммирование по ν производится, как прежде, от 1 до ∞. Поскольку здесь наивысшей степенью D в аналитической части является первая степень, то же самое должно выполняться и в правой части (37). Однако все полюсы по D предполагаются исключенными, когда мы выражаем g_{i0} в терминах g_i, поэтому $\beta_i(g, D)$ должна быть аналитична по D. Чтобы аналитическая часть выражения в правой части (37) не содержала членов выше первого порядка по D, необходимо, чтобы $\beta_i(g, D)$ была *линейна* по D:

$$\beta_i(g, D) = \beta_i^{(1)}(g)\, D + \beta_i^{(0)}(g).$$

Приравнивание членов первого и нулевого порядка по D в (41) и правой части (37) дает

$$-\rho_i g_i = \beta_i^{(1)}(g) -$$

$$-\sigma_i g_i - \sum_s b_{1i}^{(s)}(g)\,\rho_i = \beta_i^{(0)}(g) + \sum_{sj} b_{1ij}^{(s)}(g)\,\beta_j^{(1)}(g),$$

и потому

$$\beta_i(g, D) = -\rho_i g_i D - \sigma_i g_i - \sum_s b_{1i}^{(s)}(g)\rho_i + \sum_{sj} b_{1ij}^{(s)}(g)\rho_j g_j. \tag{42}$$

Тот факт, что β_i-функции линейны по D, является и замечательным и удобным; D-зависимость полностью обусловлена членом $-d_i(D)g_i$ в β_i, и петлевые вклады в (20) являются D-независимыми.

(Следует упомянуть, между прочим, что сравнение полюсных членов в (37) ведет к дополнительным соотношениям [150], которые определяют $b_{\nu i}^{(s)}$ для $\nu > 1$ в членах $b_{1i}^{(s)}$. Используя (42), можно привести правую часть (37) к виду

$$-\rho_i g_i D - \sigma_i g_i - \sum_s b_{1i}^{(s)}(g)\,\rho_i + \\ + \sum_{s,\,\nu,\,j}(D-D_s)^{-\nu}[b_{\nu ij}^{(s)}(g)\,\beta_j(g, D_s) - b_{\nu+1,\,ij}^{(s)}(g)\,\rho_j g_j.]$$

Сравнивая полюсные члены с соответствующими членами в (41), получаем рекурсивные соотношения

$$\rho_i b_{\nu+1,\, i}^{(s)}(g) - \sum_j \rho_j g_j b_{\nu+1,\, ij}^{(s)}(g) =$$

$$= -(\sigma_i + \rho_i D_s) b_{\nu i}^{(s)}(g) - \sum_j b_{\nu ij}^{(s)}(g)\, \beta_j(g, D_s) \text{ для всех } \nu \geqslant 1.) \qquad (43)$$

Дальнейшая полезная информация относительно функций вычетов может быть получена из размерных соображений. Одна из особенностей размерной регуляризации состоит в том, что полюсы в функциях Грина возникают только из логарифмических ультрафиолетовых расходимостей, а не из квадратичных или более высоких расходимостей. Однако логарифмическая расходимость может появиться в данной величине, только если размерность этой величины точно равна размерности констант связи, которым эта величина пропорциональна. Отсюда следует, что $b_i^{(s)}(g)$ может содержать член порядка $g_a g_b g_c \ldots$, только если размерность g_{i0} равна полной размерности $g_{a0} g_{b0} g_{c0} \ldots$ при пространственно-временно́й размерности D_s:

$$d_i(D_s) = d_a(D_s) + d_b(D_s) + \ldots . \qquad (44)$$

Из (42) видно, что то же самое верно для D_s-членов в β_i. Например, в евклидовой скалярной полевой теории с симметрией относительно преобразования $\phi \to -\phi$ обычный лагранжиан есть

$$\mathscr{L} = -\frac{1}{2}\partial_\mu \phi\, \partial_\mu \phi - \frac{1}{2} g_{10} \phi^2 - \frac{1}{4!} g_{20} \phi^4 - \frac{1}{6!} g_{30} \phi^6 - \frac{1}{8!} g_{40} \phi^8 \ \ldots . \qquad (45)$$

(Мы используем нашу свободу выполнять точечные преобразования с $\delta\phi$ в виде линейной комбинации ϕ, ϕ^3, $\Box\phi$, ϕ^5, $\phi^2\Box\phi$, $\Box^2\phi$ и т. д., чтобы исключить такие члены, как $\phi^3\Box\phi$, $(\Box\phi)^2$, $\phi^5\Box\phi$, $\phi^2(\Box\phi)^2$, $(\Box^2\phi)^2$ и т. д., и чтобы привести коэффициент при $-{}^1/{}_2\partial_\mu\phi\partial_\mu\phi$ к единице.) В четырех измерениях g_{i0} имеют размерности

$$d_1 = 2, \quad d_2 = 0, \quad d_3 = -2, \quad d_4 = -4, \ \ldots . \qquad (46)$$

Следовательно, если мы определим перенормированные связи так, чтобы исключить только полюсы при $D_s = 4$, то β-функции будут иметь структуру

$$\begin{gathered} \beta_1 = g_1 F_1, \quad \beta_2 = F_2, \\ \beta_3 = g_3 F_3 + g_1 g_4 F_{14} + \ldots, \\ \beta_4 = g_4 F_4 + g_3^2 F_{33} + \ldots, \end{gathered} \qquad (47)$$

где F — функции только от переменных g_2, $g_1 g_3$, $g_1^2 g_4$, ...

Этот формализм делает возможным очень компактный и удобный анализ свойств определенных фиксированных точек. Предположим, что g_i определяются так, чтобы исключить только полюсы при единственной пространственно-временно́й размерности D_s. Величина β_i для любой неперенормируемой или сверхперенормируемой связи

должна быть всегда пропорциональна первой или более высоким степеням неперенормируемой или сверхперенормируемой связей соответственно. Следовательно, набор связей g_i^* с обращающимися в нуль значениями для всех взаимодействий, которые являются неперенормируемыми или сверхперенормируемыми при D_s, будет представлять фиксированную точку лишь при условии, что величины $\beta_i(g^*)$, соответствующие перенормируемым связям, будут стремиться к нулю. Кроме того, единственными членами в $\partial\beta_i/\partial g_i$, которые не стремятся к нулю в этой фиксированной точке, являются члены с $d_i=d_j$; поэтому матрица B_{ij} диагональна по размерности связей, и собственные значения этой матрицы могут быть получены диагонализацией субматриц, соединяющих связи одинаковой размерности. Например, в теории, описываемой лагранжианом (45), все β-функции (46) будут стремиться к нулю в точке

$$g_1=0,\ g_2=g_2^*,\quad g_3=0,\ g_4=0,\ \ldots \tag{48}$$

лишь при условии, что g_2^* удовлетворяет соотношению

$$\beta_2(0,\ g_2^*,\ 0,\ \ldots)=0. \tag{49}$$

Кроме того, B-матрица для этой фиксированной точки диагональна и имеет отличные от нуля элементы:

$$B_{11}=F_1(g^*),\ B_{22}=(\partial F_2/\partial g_2)^*,\ B_{33}=F_3(g^*)\ \ldots. \tag{50}$$

Все эти диагональные элементы являются в таком случае собственными значениями B-матрицы. Вообще говоря, фиксированная точка этого типа может быть найдена с использованием строго перенормируемой теории (хотя действительная теория может и не быть перенормируемой вообще) и собственные значения B-матрицы, «критические экспоненты», могут быть получены рассмотрением всех сверхперенормируемых и неперенормируемых взаимодействий в качестве возмущений первого порядка.

Может быть, следовало бы подчеркнуть, что даже если определение (36) перенормированных связей дает тривиальную зависимость β-функций от пространственно-временно́й размерности D, критические экспоненты имеют в точности такую же сложную D-зависимость, которую они имели бы при любом другом определении перенормированных связей. Чтобы проиллюстрировать этот факт, вернемся к лагранжиану (45) с $D_s=4$. По мотивам, обсуждавшимся выше, мы можем найти фиксированную точку, используя урезанную строго перенормируемую теорию:

$$\mathcal{L}=-\frac{1}{2}\partial_\mu\phi\,\partial_\mu\phi-\frac{1}{24}\lambda_0\phi^4. \tag{51}$$

Коллинз [152] вычислил полюсы в λ_0, которые требуются для сокращения полюсов в функциях Грина при $D\to4$ снизу; в двухпетле-

вом порядке он нашел

$$\lambda_0\mu^{D-4}=\lambda+(D-4)^{-1}\left(\frac{-3\lambda^2}{16\pi^2}+\frac{17\lambda^3}{6\,(16\pi^2)^2}+\ \ldots\right)+$$
$$+(D-4)^{-2}\left(\frac{\lambda^2}{(16\pi^2)^2}+\ \ldots\right)+\ \ldots. \qquad (52)$$

Функция вычета для $\nu=1$, таким образом, имеет вид

$$b_{1\lambda}=\frac{-3\lambda^2}{16\pi^2}+\frac{17}{6}\frac{\lambda^3}{(16\pi^2)^2}+\ \ldots. \qquad (53)$$

Размерность $\sigma_\lambda+\rho_\lambda D$ величины λ_0 есть $4-D$, поэтому $\sigma_\lambda=4$, $\rho_\lambda=1$ и (42) дает здесь β-функцию в виде

$$\beta_\lambda(\lambda,\ D)=(D-4)\,\lambda+b_{1\lambda}-\lambda\,\partial b_{1\lambda}/\partial\lambda \qquad (54)$$

или с использованием (53)

$$\beta_\lambda(\lambda,\ D)=(D-4)\,\lambda+\frac{3\lambda^2}{16\pi^2}-\frac{17}{3}\frac{\lambda^3}{(16\pi^2)^2}+\ \ldots. \qquad (55)$$

Фиксированная точка, где $\beta_\lambda(\lambda^*,D)$ обращается в нуль, может быть получена в виде ряда по степеням $(D-4)$:

$$\lambda^*=16\pi^2\left|\frac{1}{3}(4-D)+\frac{17}{9}(4-D)^2+\ \ldots\ \right|. \qquad (56)$$

(Эта точка известна как фиксированная точка Вильсона — Фишера [130, 131]. Она имеет физический знак $\lambda^*>0$ только для $D<4$.) Итак, мы легко вычисляем критическую экспоненту

$$\left(\frac{\partial\beta_\lambda}{\partial\lambda}\right)_{\lambda=\lambda^*}=-(4-D)+\frac{6\lambda^*}{16\pi^2}-\frac{17\lambda^{*2}}{(16\pi^2)^2}+\ \ldots\ =(4-D)+\frac{85}{9}(4-D)^2+\ldots \qquad (57)$$

Отметим, что она положительна по крайней мере для конечной области D ниже $D=4$. Все другие критические экспоненты также положительны в окрестности $D=4$, за исключением одной, соответствующей сверхперенормируемой связи $-(^1/_2)m_0^2\phi^2$. Коллинз [152] также вычислил полюсы в m_0^2, необходимые для того, чтобы аннулировать особенности при $D=4$, вводимые этой связью; в двухпетлевом порядке его результат есть

$$m_0^2\mu^{-2}=m^2\left[1+(D-4)^{-1}\left(\frac{-\lambda}{16\pi^2}+\frac{5}{12}\frac{\lambda^2}{(16\pi^2)^2}+\ \ldots\right)+\right.$$
$$\left.+(D-4)^{-2}\frac{2\lambda^2}{(16\pi^2)^2}+\ \ldots\right]. \qquad (58)$$

Следовательно, функция вычета для $\nu=1$ и $D_s=4$ имеет вид

$$b_{1,\,m^2}=m^2\left[\frac{-\lambda}{16\pi^2}+\frac{5}{12}\frac{\lambda^2}{(16\pi^2)^2}+\ \ldots\right]. \qquad (59)$$

Размерность $\sigma+\rho D$ величины m_0^2 равна $+2$, поэтому $\sigma_{m^2}=+2$, $\rho_{m^2}=$

$=0$ и (42) дает β-функцию для m^2 в виде

$$\beta_{m^2} = -2m^2 - \lambda \frac{\partial b_{1, m^2}}{\partial \lambda} \qquad (60)$$

или с использованием (59)

$$\beta_{m^2} = m^2 \left[-2 + \frac{\lambda}{16\pi^2} - \frac{5}{6} \frac{\lambda^2}{(16\pi^2)^2} + \ \ldots \right]. \qquad (61)$$

Соответствующая критическая экспонента обычно обозначается $-\nu^{-1}$:

$$-\nu^{-1} \equiv \left(\frac{\partial \beta_{m^2}}{\partial m^2} \right)_{\lambda=\lambda^*,\ m^2=0} = -2 + \frac{\lambda^*}{16\pi^2} - \frac{5}{6} \frac{\lambda^{*2}}{(16\pi^2)^2} + \ \cdots \qquad (62)$$

или, используя (56),

$$\nu = \frac{1}{2} + \frac{1}{12}(4-D) + \frac{7}{162}(4-D)^2 + \ \ldots \qquad (63)$$

В согласии с известными результатами [153, 154]. Тот факт, что только одно собственное значение $-\nu^{-1}$ матрицы B_{ij} отрицательно для $D=4-\varepsilon$, означает, что фиксированная точка Вильсона — Фишера имеет одномерную ультрафиолетовую критическую поверхность и что есть лишь один параметр, который необходимо установить, чтобы произвести фазовый переход второго рода.

Мы видели в разд. 3, что существование теории, которая является перенормируемой и асимптотически безопасной при пространственно-временной размерности D_r, указывает на существование фиксированной точки вблизи $g^*=0$ с конечномерной критической поверхностью по крайней мере для конечной области пространственно-временных размерностей D выше D_r. Это можно очень удобно показать, используя метод продолжения по размерности, описанный в этом разделе. Предположим для простоты, что теория, которая является (строго) перенормируемой при $D=D_r$, имеет единственный параметр связи λ_0 с размерностью $-(D-D_r)\rho$, где $\rho>0$. Определим безразмерный перенормированный параметр связи $\lambda(\mu)$ таким образом, чтобы исключить все полюсы в скоростях реакций при $D=D_r$. Исходя из соображений, обсуждавшихся выше в этом разделе, мы можем найти фиксированную точку полной теории для любого D, полагая равными нулю все связи, которые были бы неперенормируемыми или сверхперенормируемыми для $D=D_r$, и находя фиксированную точку урезанной теории. Когда это сделано, уравнение ренорм-группы, которому удовлетворяет $\lambda(\mu)$, будет иметь вид

$$\mu \frac{d}{d\mu} \lambda(\mu) = \beta(\lambda(\mu),\ D) = [D - D_r]\, \rho \lambda(\mu) + \beta(\lambda(\mu),\ D_r). \qquad (64)$$

Второй член справа возникает из петлевых диаграмм, поэтому его степенной ряд будет, вообще говоря, начинаться с членов второго

порядка:

$$\beta(\lambda(\mu),\ D_r) = -b\lambda^2(\mu) + O(\lambda^3(\mu)). \tag{65}$$

Чтобы теория была асимптотически свободной при $D=D_r$ (и $\lambda(\mu)>0$), необходимо, чтобы величина b была *положительной*. Тогда при положительном и достаточно малом $(D-D_r)$ будет существовать фиксированная точка

$$\lambda^* = (D-D_r)\,\rho/b + O((D-D_r)^2). \tag{66}$$

Все критические экспоненты положительны, за исключением одной, связанной с λ,

$$\left(\frac{\partial\beta_\lambda}{\partial\lambda}\right)_{\lambda=\lambda^*} = (D-D_r)\rho - 2b\lambda^* + O(\lambda^{*2}) = -(D-D_r)\rho + O((D-D_r)^2) < 0, \tag{67}$$

и тех, которые связаны с любыми массами или такими связями, которые были бы сверхперенормируемыми при $D=D_r$. Поэтому ультрафиолетовая критическая поверхность является конечномерной, состоящей как раз из тех теорий, которые были бы перенормируемыми при $D=D_r$.

Это не следует интерпретировать как утверждение о том, что эти асимптотически безопасные теории являются перенормируемыми, как обычно для $D>D_r$. Метод размерной регуляризации здесь несколько вводит в заблуждение — он устраняет ультрафиолетовые расходимости во *всех* теориях при нерациональных значениях пространственно-временно́й размерности D ценой введения полюсов при рациональных значениях D. При любой другой схеме регуляризации имеется множество ультрафиолетовых расходимостей при $D>D_r$, и они должны быть устранены включением в лагранжиан всех возможных взаимодействий, допускаемых симметриями теории. Появление фиксированной точки (66) с конечномерной критической поверхностью в формализме размерной регуляризации гарантирует, что *существует* фиксированная точка с ультрафиолетовой критической поверхностью такой же размерности в более традиционных схемах перенормировки, но фиксированная точка там будет иметь, вообще говоря, не обращающиеся в нуль значения для *всех* связей, перенормируемых *и* неперенормируемых, и асимптотически безопасные теории не будут даже казаться перенормируемыми в обычном смысле.

Почему же в таком случае мы должны отказаться от формализма размерной регуляризации, в котором асимптотически безопасные теории оказываются столь простыми? Причина заключается как раз в том, что в конечном счете мы должны заниматься физической пространственно-временно́й размерностью $D=4$, которая больше, чем D_r, и при продолжении от $D=D_r$ до $D=4$ мы должны избегать полюсов при промежуточных рациональных значениях D и при самом $D=4$, которые будут присутствовать при размерной регуляри-

зации. Обычная схема перенормировки предлагает возможный способ осуществления этого продолжения ценой отказа от видимой перенормируемости. Однако первый шаг к тому, чтобы убедиться, что существует фиксированная точка с конечномерной критической поверхностью для D именно выше D_r, является тем шагом, который может быть легче всего выполнен методом размерной регуляризации.

6. ГРАВИТАЦИЯ В 2+ε-ИЗМЕРЕНИЯХ

Вернемся наконец к гравитации. Мы хотим знать, можно ли требовать, чтобы квантовая теория гравитации была асимптотически безопасной, и сколько свободных параметров имелось бы в такой теории. Это зависит от того, существует ли фиксированная точка g^*, и от размерности ее критической поверхности.

Чтобы подойти к этому вопросу, мы используем метод продолжения по размерности, обсуждавшийся в предыдущем разделе. В двух измерениях существует единственная, строго перенормируемая теория чистой гравитации, основанная на эйнштейновском лагранжиане $-\sqrt{g}R/16\pi G$. (Интеграл $\int d^2x\sqrt{g}R$ является безразмерным в двумерном пространстве, так что величина G должна быть безразмерной, чтобы действие $\int d^2x\mathscr{L}$ не было размерным.) Теория остается перенормируемой, если мы добавляем к гравитации материальные поля с минимальной связью, хотя в таком случае может оказаться необходимым добавить связи материальных полей друг с другом.

Конечно, общая теория относительности отнюдь не является теорией в двух измерениях. Лагранжиан $\sqrt{g}R$ есть полная производная для $D=2$, и вследствие этого левая часть полевых уравнений Эйнштейна $R_{\mu\nu}-\frac{1}{2}g_{\mu\nu}R$ обращается в нуль тождественно (см., например, [20]). Это создает проблему при использовании метода размерной регуляризации. Предположим, что когда мы вычисляем инвариантные амплитуды в 2+ε-измерениях, мы находим, что какая-то инвариантная амплитуда имеет полюс при $\varepsilon=0$ и что для того, чтобы аннулировать этот полюс, мы должны добавить к лагранжиану член, пропорциональный $\sqrt{g}R/\varepsilon$. Можем ли мы игнорировать такие контрчлены на основании того, что $\int d^2x\sqrt{g}R$ обращается в нуль при $\varepsilon=0$? Ответ оказывается отрицательным [155]: если мы не включили контрчлен, пропорциональный $\sqrt{g}R/\varepsilon$, туда, где нужно аннулировать полюсы в инвариантных амплитудах при $\varepsilon=0$, то функции Грина, которые мы вычисляем, могли бы оказаться конечными при $\varepsilon=0$, но они не будут аналитическими функциями ε при $\varepsilon=0$ в 2+ε-измерениях, как предполагалось в формулировании продолжения по размерности ренорм-группы. В таком случае мы

приходим к выводу, что гравитационная константа связи, которая появляется в эйнштейновском лагранжиане $-\sqrt{g}R/16\pi G$, подлежит перенормировке в $2+\varepsilon$-измерениях даже для $\varepsilon\to 0$. Позднее мы вернемся к вопросу о том, является ли G *существенной* связью, которую нельзя изменить подходящим переопределением полей.

Неперенормированная гравитационная постоянная $G_0(\varepsilon)$ в $2+\varepsilon$-измерениях имеет размерность $[\text{масса}]^{-\varepsilon}$, поэтому в обозначениях разд. 5 $\rho_G=-1$, $\sigma_G=0$. Структура особенности $G_0(\varepsilon)$ при $\varepsilon\to 0$ создается здесь соотношением (36) в виде

$$G_0(\varepsilon)\,\mu^{\varepsilon}\longrightarrow G(\mu)+\sum_{\nu=1}^{\infty}\varepsilon^{-\nu}b_\nu(G(\mu)). \tag{68}$$

Кроме того, (38) и (42) дают уравнение ренорм-группы для *конечного* значения ε

$$\mu\frac{d}{d\mu}G(\mu)=\beta(G(\mu),\ \varepsilon), \tag{69}$$

где

$$\beta(G,\ \varepsilon)=\varepsilon G+b_1(G)-Gb_1'(G). \tag{70}$$

Для малых G мы ожидаем, что

$$b_1(G)=bG^2+O(G^3), \tag{71}$$

так что (70) дает

$$\beta(G,\ \varepsilon)=\varepsilon G-bG^2+O(G^3). \tag{72}$$

Ключевой вопрос состоит в том, является ли b положительной величиной; и тогда существует фиксированная точка

$$G^*=\varepsilon/b+O(\varepsilon^2), \tag{73}$$

и, как показано в разд. 5, она имеет ультрафиолетовую критическую поверхность конечной размерности.

Вычисление b было выполнено в ряде конкретных случаев несколькими различными группами [1]). Их результаты, полученные к 1977 г., можно обобщить в виде утверждения, что особенности во всех чисто гравитационных функциях Грина в $2+\varepsilon$-измерениях при $\varepsilon=0$ погашаются в однопетлевом порядке, если мы предположим, что голая гравитационная постоянная имеет полюс

$$G_0\mu^{\varepsilon}\longrightarrow G+bG^2/\varepsilon, \tag{74}$$

где

$$b=\frac{38}{3}+4N_V-\frac{1}{3}N_F-\frac{2}{3}N_S. \tag{75}$$

Четыре члена возникают здесь из однопетлевых диаграмм, внутренние линии которых являются соответственно либо гравитонными линиями, либо материальными линиями спина 1, $^1/_2$ или 0;

[1]) Вклады спина 0 см. в работе [156], вклады спина $^1/_2$ — в работе [157], вклады спина 1 — в работе [158], вклады гравитона — в работе [159].

N_V и N_S — число действительных векторных и скалярных полей, а N_F — число майорановских фермионных полей. Поэтому на первый взгляд кажется, что $b>0$ и, следовательно, общая теория относительности является асимптотически безопасной в $2+\varepsilon$-измерениях при условии, что имеется достаточно калибровочных полей, чтобы сбалансировать любые скалярные или фермионные поля.

Однако прежде чем сделать какое-либо заключение, мы должны обратить особое внимание на физическую интерпретацию соотношения (75). В чистой общей теории относительности след вакуумных полевых уравнений Эйнштейна дает $R=0$ для любой пространственно-временно́й размерности, поэтому, как объяснено в разд. 3, коэффициент $1/16\pi G$ этого лагранжиана не является существенной связью, и нет никакой причины, по которой надо было бы требовать достижения фиксированной точки при $\mu\to\infty$. То же самое верно, если мы добавляем к теории любое число «фотонных» полей с чисто гравитационными взаимодействиями; в этом случае полевые уравнения Эйнштейна дают член $\sqrt{g}R$, пропорциональный $\sqrt{g}\Sigma F_{\mu\nu}F^{\mu\nu}$, но уравнения Максвелла позволяют нам переписать это в виде полной производной $\partial_\mu[\sqrt{g}\Sigma A_\nu F^{\mu\nu}]$, поэтому здесь $\partial\mathscr{L}/\partial G$ оказывается полной производной и G снова не является существенной связью. (Простейшая теория, которая перенормируема в двух измерениях и которая *действительно* имеет существенные связи,— это теория Эйнштейна — Янга — Миллса [160]. В этом случае скорости реакций зависят от единственной существенной связи $e^{2\varepsilon}G^{2-\varepsilon}$, где e — калибровочная константа связи. Однако эта величина является безразмерной для всех ε и поэтому удовлетворяет тривиальному ренорм-групповому уравнению $d(e^{2\varepsilon}G^{2-\varepsilon})/d\mu=0$.)

Недавно новую интерпретацию этих вычислений предложили Гастманс, Каллош и Труффин [161] (см. также [127]) (ГКТ). Их исходный пункт состоит в пересмотре структуры лагранжиана общей теории относительности. Гиббонс и Хокинг [162, 163] [1]) подчеркивали, что, применяя функциональный формализм к общей теории относительности, мы используем в действительности не эйнштейновский лагранжиан $-\sqrt{g}R/16\pi G_0$, а скорее

$$\mathscr{L}_G=-\frac{1}{16\pi G_0}[\sqrt{g}R-\Phi], \tag{76}$$

где Φ — полная производная, составленная так, чтобы $\mathscr{L}_G$ была функцией только от $g_{\mu\nu}$ и ее *первых* производных:

$$\Phi=\frac{\partial}{\partial x^\mu}\left\{\frac{\sqrt{g}}{2}\left[g^{\lambda\nu}g^{\mu\varkappa}\frac{\partial g_{\lambda\nu}}{\partial x^\varkappa}-g^{\lambda\nu}g^{\mu\varkappa}\frac{\partial g_{\varkappa\nu}}{\partial x^\lambda}-g^{\lambda\nu}g^{\mu\varkappa}\frac{\partial g_{\lambda\varkappa}}{\partial x^\nu}+g^{\lambda\mu}g^{\nu\varkappa}\frac{\partial g_{\nu\varkappa}}{\partial x^\lambda}\right]\right\}. \tag{77}$$

Конечно, добавление полной производной не играет роли, если мы

[1]) В работе [166] Кристенсен и Дафф связывают дополнительный член в (76) с более ранней работой Арновитта, Дезера и Мизнера [164, 165].

ограничиваем наше внимание метриками, которые достаточно быстро убывают при $|X|\to\infty$, но в функциональной формулировке квантовой гравитации мы должны суммировать по *всем* метрикам в евклидовом пространстве-времени, и наличие Ф делает важным вклад в действие для некоторых из этих метрик, таких, как евклидова метрика Шварцшильда.

Теперь, когда мы вычислили однопетлевые диаграммы в $2+\varepsilon$-измерениях, можно ожидать найти полюсы $1/\varepsilon$ при $\varepsilon\to 0$, которые требуют независимых контрчленов, пропорциональных как $\sqrt{g}R$, так и Ф. ГКТ доказывают, что поэтому лагранжиан должен записываться в таком случае в виде суммы двух независимых членов:

$$\mathcal{L}_G=-\frac{1}{16\pi G_0}[\sqrt{g}R-\Phi]-\frac{1}{16\pi F_0}\sqrt{g}R. \qquad (78)$$

Используя след полевых уравнений Эйнштейна, мы можем выразить $\sqrt{g}R$ в терминах материальных полей, так что его коэффициент не является независимой существенной связью. Однако гравитационная связь G_0 появляется в коэффициенте отличного от него члена $\sqrt{g}R-\Phi$, который *не* задается полевыми уравнениями в терминах материальных полей, поэтому G_0 есть независимая существенная связь. Другими словами, более ранние вычисления [156—159] дали правильные контрчлены в $1/G_0+1/F_0$, однако физически интересной существенной связью является $1/G_0$, и ее контрчлены должны быть вычислены заново.

ГКТ вычислили полюсы в G_0, требуемые для компенсации $1/\varepsilon$ полюсов в гравитационных функциях Грина в теории с N_S и N_V действительными скалярными и векторными полями и N_F и N_Δ майорановскими полями спинов $^1/_2$ и $^3/_2$. Их результаты можно суммировать как утверждение, что G_0 должна иметь полюс (74), где b теперь определяется так:

$$b=\frac{2}{3}\left[1+\frac{15}{2}N_\Delta-N_F-N_S\right], \qquad (79)$$

причем первый член появляется из гравитонных петель. Недавно Кристенсен и Дафф [166] (КД) выполнили вычисление в том же направлении и нашли формулу для b с другими фермионными вкладами:

$$b=\frac{2}{3}[1-N_\Delta+N_F-N_S]. \qquad (80)$$

В этом и другом случае существует асимптотически безопасная теория чистой гравитации в $2+\varepsilon$-измерениях с одномерной критической поверхностью. Асимптотическая безопасность сохраняется также, когда мы добавляем материальные поля, при условии что мы добавляем поля спина $^3/_2$ (ГКТ) или $^1/_2$ (КД), чтобы сбалансировать вклады полей спинов нуль и $^1/_2$ (ГКТ) или $^3/_2$ (КД), и при условии

также, что связи материальных полей между собой не создают проблем.

Можно заметить, что выражения (79) или (80) и (75) дают одинаковые результаты для вклада в b скалярных частиц, так что в этом случае более ранние вычисления [156—159] фактически дали контрчлены, пропорциональные $\sqrt{g}R-\Phi$, а не $\sqrt{g}R$. Согласно и ГКТ, и КД, новая характерная черта, вводимая различием между $\sqrt{g}R-\Phi$ и $\sqrt{g}R$, заключается в том, что вклад в b частиц произвольного спина просто пропорционален числу степеней свободы их полей, но (согласно КД) с дополнительным знаком «минус» для фермионов. (Множитель $^{15}/_{2}$ в (79) здесь вносит путаницу.) Например, симметричное бесшпуровое тензорное поле имеет $^{1}/_{2}\,D(D+1)-1$ независимых компонент, D из которых исключается условием калибровки, определяющим $\partial_\mu h^\mu_\nu$, и $D-1$ устраняются произволом в выполнении дополнительных калибровочных преобразований $\delta h_{\mu\nu}=\partial_\mu\phi_\nu+\partial_\nu\phi_\mu$, где $\partial_\mu\phi^\mu=0$ и $\Box\phi^\mu=0$, следовательно, число степеней свободы гравитационного поля в D измерениях есть

$$\frac{1}{2}D(D+1)-1-D-(D-1)=\frac{1}{2}D(D-3).$$

Оно равно — 1 для $D=2$, поэтому вклад гравитона в b должен быть равным и противоположным по знаку вкладу одной бесспиновой частицы. С другой стороны, векторное поле A_μ имеет D компонент, одна из которых устраняется условием калибровки, определяющим $\partial_\mu A^\mu$, а другая — произволом в выполнении дополнительных калибровочных преобразований $\delta A_\mu=\partial_\mu\phi$, где $\Box\phi=0$; следовательно, число степеней свободы фотонного поля в D измерениях есть $D-2$. Оно обращается в нуль при $D=2$, поэтому фотоны не дают вклада в b.

Поскольку вычисление вклада скалярных полей в b, очевидно, не создает проблем, связанных с различием между $\sqrt{g}R$ и $\sqrt{g}R-\phi$, и поскольку этот вклад устанавливает масштаб для вкладов частиц не нулевого спина, интересно было бы провести несколько более подробное вычисление этой величины. Расчет, основанный на методах [76, 77], представлен в приложении.

Интересно применить эти результаты к расширенным теориям супергравитации [1]), в которых гравитон появляется в мультиплете с полями более низкого спина. В четырехмерной теории с $n\leqslant 7$ генераторами суперсимметрии гравитон спиральности $+2$ будет появляться в мультиплете с $\left[{n \atop r}\right]$ безмассовыми частицами спиральности $2-r/2$, где $r=1, 2, \ldots, n$, и имеется отдельный мультиплет, содержащий -2-гравитон и $\left[{n \atop r}\right]$ безмассовых частиц спиральности $-2+r/2$. Для $n=8$ существует единственный мультиплет, содержащий гравитоны спиральности ± 2 и $\left[{8 \atop r}\right]$ безмассовых частиц спиральности $2-r/2$. Мы предполагаем, что для произвольной про-

[1]) См. примечание [2]) на с. 411.

странственно-временно́й размерности число полей данного спина S равно числу полей этого спина в четырех измерениях и, следовательно, числу состояний в четырех измерениях со спиральностью $+S$ (или $-S$, но не обеих сразу), даже если суперсимметрия фактически присутствует только для четырех измерений. Числа N_Δ, N_V, N_F, N_S полей спина $^3/_2$ 1, $^1/_2$, 0 в $O(n)$-расширенной супергравитации представлены в табл. 1 вместе со значениями b, полученными из (79) или (80). Согласно ГКТ-результатам, $b>0$ для чистой супергравитации ($n=1$), а также для $O(n)$-расширенной супергравитации с $n\leqslant 5$. Однако если результаты КД справедливы, то требуется двухпетлевой расчет для того, чтобы решить вопрос об асимптотической безопасности в чистой супергравитации, хотя мы всегда будем иметь $b>0$ для $n=1$, если добавить достаточно «векторных» супермультиплетов со спинами 1 и $^1/_2$. Согласно КД-результату, оказывается невозможным иметь $b>0$ в $O(n)$-расширенной супергравитации с $n\geqslant 2$ независимо от того, добавляем ли мы дополнительные материальные супермультиплеты.

Таблица 1

Числа типов полей и значения b в расширенных суперсимметричных теориях с n генераторами суперсимметрии

n	N_Δ	N_V	N_F	N_S	$3b$ (ГКТ)	$3b$ (КД)
0	0	0	0	0	2	2
1	1	0	0	0	17	0
2	2	1	0	0	32	—2
3	3	3	1	0	45	—2
4	4	6	4	2	50	—2
5	5	10	11	10	35	—6
6	6	16	26	30	—20	—18
7	8	28	56	70	—130	—42
8	8	28	56	70	—130	—42

Практически важным здесь является вопрос о продолжении на четыре измерения. В этом вопросе формализм размерной регуляризации в $2+\varepsilon$-измерениях может несколько вводить в заблуждение. Верно, что асимптотически безопасная теория гравитации в этом формализме основывается на лагранжианах, которые должны быть перенормируемыми в двух измерениях, и эти лагранжианы не содержат каких-либо контрчленов, которые компенсировали бы полюсы в фейнмановских диаграммах при пространственно-временно́й размерности $D=4$. Однако наличие этих полюсов указывает, что разложение в ряд теории возмущений по степеням ε потеряет смысл задолго до того, как мы достигнем $\varepsilon=2$.

Более благоприятная картина возможностей продолжения от $2+\varepsilon$ к четырем измерениям может быть получена в обычной перенормировочной схеме, с которой мы начинали в разд. 3. В этом формализме интегралы регулируются некоторой процедурой ультрафиолетового обрезания Λ, и зависимость от обрезания при $\Lambda\to\infty$ устраняется сокращением с контрчленами, которые дают неперенормированные константы связи. Имеется бесконечное число контрчленов, возникающих при $D>2$, и такая теория гравитации не может называться перенормируемой в обычном смысле, однако не имеется также никаких новых особенностей, появляющихся при приближении D к четырем. Один результат метода размерной регуляризации, который может быть прямо перенесен в обычную перенормировочную схему, заключается в том, что в $2+\varepsilon$-измерениях для достаточно малого ε существует фиксированная точка с ультрафиолетовой критической поверхностью конечной размерности.

БЛАГОДАРНОСТИ

Мне приятно поблагодарить С. Коулмена, С. Дезера и М. Дж. Даффа за их частую и полезную помощь в подготовке этой статьи. Я также признателен Л. Брауну, С. Кристенсену, Дж. К. Коллинзу, Б. Де Витту, Е. С. Фрадкину, Г. т'Хофту, Б. Ли, П. Мартину, Д. Нельсону, А. Саламу, Х. Шнитцеру, Л. Смоллину, Х. С. Цао, П. ван Ньювенхьюзену, К. Вильсону, Е. Виттену и Б. Зумино за информативные обсуждения различных конкретных проблем.

7. ПРИЛОЖЕНИЕ. ВЫЧИСЛЕНИЕ b

В этом приложении дается расчет вычета полюса с «голой» гравитационной постоянной при $D=2$, полученной с использованием метода работ [76, 77].

Введем гравитационное поле $h_{\mu\nu}$, такое, что

$$g_{\mu\nu}=\eta_{\mu\nu}+(32\pi G)^{1/2}h_{\mu\nu}, \qquad (\text{П}.1)$$

и будем работать в калибровке, в которой

$$\partial_\mu h^{\mu\nu}=0. \qquad (\text{П}.2)$$

(Индексы здесь поднимаются и опускаются с помощью $\eta_{\mu\nu}$, а не $g_{\mu\nu}$.) «Голый» гравитонный пропагатор пространственно-временной размерности D есть тогда

$$\langle T\{h_{\mu\nu}(x),\ h_{\lambda\rho}(0)\}\rangle_0=\int\frac{d^Dq}{(2\pi)^D}e^{iq\cdot x}\Delta_{\mu\nu,\,\lambda\rho}(q), \qquad (\text{П}.3)$$

$$\Delta_{\mu\nu,\,\lambda\rho}(q)=\frac{1}{2q^2}\left[L_{\mu\rho}(q)\,L_{\nu\lambda}(q)+L_{\mu\lambda}(q)\,L_{\nu\rho}(q)-\frac{2}{D-2}L_{\mu\nu}(q)\,L_{\lambda\rho}(q)\right]; \qquad (\text{П}.4)$$

$$L_{\mu\nu}(q)\equiv\eta_{\mu\nu}-q_\mu q_\nu/q^2. \qquad (\text{П}.5)$$

В однопетлевом порядке этот пропагатор принимает вид

$$\Delta'_{\mu\nu,\lambda\rho}(q) = \Delta_{\mu\nu,\lambda\rho}(q) + \Delta_{\mu\nu,\mu'\nu'}(q)\,\Pi^{\mu'\nu',\lambda'\rho'}(q)\,\Delta_{\lambda'\rho',\lambda\rho}(q), \quad (\Pi.6)$$

где Π — тензор гравитонной поляризации вакуума, определяемый как

$$\langle T\{T^{\mu\nu}(x), T^{\lambda\rho}(0)\}\rangle_0 = \frac{-i}{8\pi G}\int \frac{d^Dq}{(2\pi)^D} e^{iq\cdot x}\Pi^{\mu\nu,\lambda\rho}(q). \quad (\Pi.7)$$

Так как тензор энергии-импульса сохраняется, Π можно записать как

$$\Pi^{\mu\nu,\lambda\rho}(q) = (q^2)^2 A(q^2) L^{\mu\nu}(q) L^{\lambda\rho}(q) - \\ - q^2 B(q^2)\left[L^{\mu\lambda}(q) L^{\nu\rho}(q) + L^{\mu\rho}(q) L^{\nu\lambda}(q) - 2L^{\mu\nu}(q) L^{\lambda\rho}(q)\right], \quad (\Pi.8)$$

где $A(q^2)$ и $B(q^2)$ свободны от полюсов при $q^2=0$. Теперь можно прямо вычислить исправленный пропагатор как

$$\Delta'^{\mu\nu,\lambda\rho}(q) = \frac{1}{2q^2}(1-2B(q^2))\Big[L^{\mu\rho}(q) L^{\mu\lambda}(q) + \\ + L^{\mu\lambda}(q) L^{\nu\rho}(q) - \frac{2}{D-2} L^{\mu\nu}(q) L^{\lambda\rho}(q)\Big] + \frac{A(q^2)}{(D-2)^2} L^{\mu\nu}(q) L^{\lambda\rho}(q). \quad (\Pi.9)$$

Мы видим, что перенормированная гравитационная постоянная, которая определяет силу дальнодействующего гравитонного обмена, есть

$$G = G_0(1-2B(0)), \quad (\Pi.10)$$

где G_0— «голая» гравитационная постоянная. Следовательно, если $B(0)$ имеет полюс при $D=2$ вида

$$B(0, D) \to \frac{bG}{2}\left(\frac{1}{D-2}\right), \quad (\Pi.11)$$

мы должны ввести в G_0 полюс вида

$$\mu^{D-2}G_0 \underset{D\to 2}{\to} G + bG^2\left(\frac{1}{D-2}\right). \quad (\Pi.12)$$

Таким образом, величина b в (П.11) является такой же, как в (71) — (73). Наша задача заключается в вычислении вычета полюса в $B(q^2, D)$ при $D=2$ и в определении b сравнением с (П. 11).

Для этой цели мы введем спектральные функции ρ_A, ρ_B по формуле

$$\sum \delta^D(p_n - p)\langle 0|T^{\mu\nu}(0)|n\rangle \langle 0|T^{\rho\sigma}(0)|n\rangle^* = \\ = (2\pi)^{-D+1}\theta(p^0)\left[\rho_A(-p^2)(p^2)^2 L^{\mu\nu}(p) L^{\rho\sigma}(p) - \\ - p^2\rho_B(-p^2)\{L^{\mu\rho}(p) L^{\nu\sigma}(p) + L^{\mu\sigma}(p) L^{\nu\rho}(p) - 2L^{\mu\nu}(p) L^{\mu\rho}(p)\}\right]. \quad (\Pi.13)$$

Не учитывая возможных вычитаний, из (П.7), (П.8) и (П.13) полу-

чаем

$$A(q^2) = 8\pi G \int_0^\infty [q^2 + \mu^2 - i\varepsilon]^{-1} \rho_A(\mu^2)\, d\mu^2, \qquad (\text{П}.14)$$

$$B(q^2) = 8\pi G \int_0^\infty [q^2 + \mu^2 - i\varepsilon]^{-1} \rho_B(\mu^2)\, d\mu^2. \qquad (\text{П}.15)$$

Рассмотрим теперь вклад состояния, состоящего из пары идентичных нейтральных бесспиновых частиц массы m и импульсов $\vec{k}$, $\vec{k}'$, в ρ_A и ρ_B. В наинизшем порядке теории возмущений для любого D [1])

$$\langle 0 | T^{\mu\nu}(0) | \vec{k}, \vec{k}' \rangle = -(2\pi)^{-(D-1)} (2\omega)^{-1/2} (2\omega')^{-1/2} \times$$
$$\times \Big[k^\mu k'^\nu + k'^\mu k^\nu + \eta^{\mu\nu}(-k^\lambda k'_\lambda + m^2) -$$
$$- \frac{(D-2)}{2(D-1)} \{(k+k')^\mu (k+k')^\nu - (k+k')^2 \eta^{\mu\nu}\} \Big], \qquad (\text{П}.16)$$

где

$$k^0 = \omega = (\vec{k}^2 + m^2)^{1/2}, \quad k^{0\prime} = \omega' = (\vec{k}'^2 + m^2)^{1/2}.$$

Уравнение (П.13) тогда дает

$$\rho_A(\mu^2) = \frac{1}{8} (2\pi)^{-D+1} \Omega_D k^{D-3} \mu^{-5} \left[\frac{12k^4}{D^2-1} - \frac{2k^2\mu^2}{(D-1)^2} + \frac{\mu^4}{4(D-1)^2} \right], \qquad (\text{П}.17)$$

$$\rho_B(\mu^2) = \frac{1}{2} (2\pi)^{-D+1} \Omega_D k^{D+1} \mu^{-3} / (D^2 - 1), \qquad (\text{П}.18)$$

где Ω_D — площадь поверхности единичной сферы в $(D-1)$ пространственных измерениях

$$\Omega_D = 2\pi^{(D-1)/2} / \Gamma\left(\frac{D-1}{2}\right),$$

$$k \equiv \left(\frac{\mu^2}{4} - m^2\right)^{1/2}.$$

Теперь мы можем вычислить b. Функция $B(q^2, D)$ задается формулами (П.15) и (П.18) как

$$B(q^2, D) = \frac{4\pi G \Omega_D}{(D^2-1)(2\pi)^{D-1}} \int_{4m^2}^\infty \frac{k^{D+1} d\mu^2}{\mu^3 (\mu^2 + q^2)}. \qquad (\text{П}.19)$$

Этот интеграл является хорошо определенным для $D<2$ и может быть аналитически продолжен до $D>2$ с полюсом при $D=2$:

$$B(q^2, D) \xrightarrow[D\to 2]{} \frac{G}{3} \left(\frac{1}{2-D}\right). \qquad (\text{П}.20)$$

---

[1]) В этой формуле подразумевается, что вид тензора энергии-импульса «усовершенствован» для произвольного D способом, предложенным в работе [167], однако это влияет лишь на $A(q^2)$, но не на $B(q^2)$.

Сравнивая с (П.11), мы видим, что

$$b=-2/3 \qquad (\Pi.21)$$

в согласии с (79) и (80) при $N_S=1$.

Этот метод расчета имеет то преимущество, что позволяет нам делать выводы общего характера относительно знака и других свойств спектральных функций и вычетов при различных размерностях для промежуточных состояний произвольного спина. Примем, что p в (П.13) лежит во «временно́м» направлении $p=(0, \ldots, 0, \mu)$, и свернем с $a_\mu b_\nu a_\rho b_\sigma$, где a и b — чисто пространственные векторы, разделенные углом ϕ; это дает

$$(1+\operatorname{tg}\phi)^2 \rho_A(\mu^2)\mu^4 - 4\operatorname{tg}\phi\, \rho_B(\mu^2)\mu^2 \geqslant 0 \qquad (\Pi.22)$$

для всех ϕ и μ. Для $D=2$ мы должны, конечно, принять $\phi=0$, так что это дает только условие

$$\rho_A(\mu^2) \geqslant 0. \qquad (\Pi.23)$$

Для целых размерностей $D \geqslant 3$ можно брать ϕ произвольно; выбирая его так, чтобы минимизировать левую часть (П.22), найдем

$$0 \leqslant \rho_B(\mu^2) \leqslant \rho_A(\mu^2)\mu^2. \qquad (\Pi.24)$$

Наконец, для бесшпурового тензора энергии-импульса (П.13) дает

$$0=\rho_A(\mu^2)\mu^2(D-1)+\rho_B(\mu^2)(4-2D). \qquad (\Pi.25)$$

В частности, $\rho_A(\mu^2)=0$ для $D=2$. Даже если конечные массы дают тензор энергии-импульса с ненулевым следом, (П.25) будет асимптотически справедливо при $\mu \to \infty$, поэтому интеграл (П.14) для $A(q^2)$ не будет иметь полюса при $D=2$.

ЛИТЕРАТУРА

1. *Heisenberg W.*, Z. Physik, **110**, 251 (1938).
2. *Heisenberg W.*, Physik, **101**, 251 (1938); Z. Physik, **113**, 61 (1939).
3. *Sakata S.*, *Umezawava H.*, *Kamefuchi S.*, Progr. Theor. Phys., **7**, 327 (1952).
4. *Feynman R. P.*, Acta Phys. Polon., **24**, 697 (1963).
5. *DeWitt B. S.*, Phys. Rev., **162**, 1195, 1239 (1967); erratum Phys. Rev., **171**, 1834 (1968).
6. *Faddeev L. D.*, *Popov V. N.*, Phys. Lett., **B25**, 29 (1967).
7. *Mandelstam S.*, Phys. Rev., **175**, 1604 (1968).
8. *Fradkin E. S.*, *Tyutin J. V.*, Phys. Rev., 2, 2841 (1970).
9. *Fradkin E. S.*, *Vilkowsky G. A.*, Phys. Lett., **55B**, 224 (1975); Nuovo Cimento, **13**, 187 (1975).
10. *Duff M. J.*, In: Quantum gravity, eds. Isham C. J., Penrose, Sciama D. W., Oxford Univ. Press, 1975.
11. *'t Hooft G.*, Nucl. Phys., **B62**, 444 (1973).
12. *'t Hooft G.*, *Veltman M.*, Ann. Inst. Poincaré, **20**, 69 (1974).
13. *Deser S.*, *van Nieuwenhuizen P.*, Phys. Rev. Lett., **32**, 245 (1974).
14. *Deser S.*, *van Nieuwenhuizen P.*, Phys. Rev., **D10**, 401, 411 (1974).
15. *Deser S.*, *Tsao H.-S.*, *van Nieuwenhuuizen P.*, Phys. Lett., **50B**, 491 (1974).
16. *Weinberg S.*, Phys. Rev., **140**, B546 (1965).
17. *Utiyama R.*, *DeWitt B. S.*, J. Math. Phys., **3**, 608 (1962).

18. *Georgi H., Quinn H., Weinberg S.*, Phys. Rev. Lett., **33**, 451 (1974).
19. *Planck M.*, Sitz. Deut. Akad. Wiss. (Berlin), 440 (1899).
20. *Weinberg S.*, Gravitation and Cosmology—Principles and Applications of the General Theory of Relativity, Wiley, New York, 1972. [Имеется перевод: *Вейнберг С.* Гравитация и космология.— М.: Мир, 1975.]
21. *Gell-Mann M., Low F. E.*, Phys. Rev., **95**, 1300 (1954).
22. *Stueckelberg E. C. G., Petermann A.*, Helv. Phys. Acta, **26**, 499 (1953).
23. *Wilson K. G., Kogut J.*, Phys. Rep., **12C**, No. 2 (1974).
24. *Fisher M. E.*, Rev. Mod. Phys., **46**, 597 (1974).
25. *Brézin E., LeGuillou J. C., Zinn-Justin J.*, In: Phase Transition and Critical Phenomena, eds. C. Domb and M. S. Green, Academic Press, London, New York, 1975.
26. *Wegner F. J.*, In: Trends in Elementary Particle Theory, p. 171, Springer-Verlag, Berlin, 1975.
27. *Wilson K.*, Rev. Mod. Phys., **47**, 773 (1975).
28. *Shand-Keng Ma.*, Modern Theory of Critical Phenomena, W. A. Benjamin, New York, 1976.
29. *Wess J., Zumino B.*, Nucl. Phys., **B70**, 34 (1974).
30. *Wess J., Zumino B.*, Nucl. Phys., **B78**, 1 (1974).
31. *Wess J., Zumino B.*, Phys. Lett., **49B**, 52 (1974).
32. *Гольфанд Ю. А., Лихтман Е. П.*, Письма в ЖЭТФ, **13**, 452 (1971).
33. *Volkov D. V., Akulov V. P.*, Phys. Lett., **46B**. 109 (1973).
34. *Salam A., Strathdee J.*, Nucl. Phys., **B76**, 477 (1974).
35. *Fayet P., Ferrara S.*, Phys. Rep., **32C**, 249 (1977).
36. *Freedman D. Z., van Nieuwenhuizen P., Ferrara S.*, Phys. Rev., **D13**, 3214 (1976).
37. *Deser S., Zumino B.*, Phys. Lett., **62B**, 335 (1976).
38. *Zumino B.*, CERN preprint, 1977.
39. *Deser S.*, Brandeis preprint, 1978.
40. *Freedman D. Z., van Nieuwenhuizen P.*, Rev. Mod. Phys. (1978).
41. *Kaku M., Townsend P. K., van Nieuwenhuizen P.*, Phys. Rev. Lett., **39**, 1109 (1977).
42. *Ferrara S., van Nieuwenhuizen P.*, Ecole Normale Superieur preprint 78/14, 1978.
43. *Wess J., Zumino B.*, Phys. Lett., **49B**, 52 (1974).
44. *Iliopoulos J., Zumino B.*, Nucl. Phys., **B76**, 310 (1974).
45. *Ferrara S., Iliopoulos J., Zumino B.*, Nucl. Phys., **B77**, 413 (1974).
46. *Zumino B.*, Nucl. Phys., **B89**, 535 (1975).
47. *Zumino B.*, In: Proceedings of the XVIII International Conference on High Energy Physics, ed. J. R. Smith Rutherford Laboratory, Chilton, Didcot, Oxfordshire, 1974.
48. *Freedman D. Z., van Niewenhuizen P., Ferrara S.*, Phys. Rev. **D13**, 3214 (1976); Phys. Rev., **D14**, 912 (1976).
49. *Deser S., Zumino B.*, Phys. Lett., **62B**, 335 (1976).
50. *Grisaru M. T., van Niewenhuizen P., Vermaseren J. A. M.*, Phys. Rev. Lett., **37**, 1662 (1976).
51. *Deser S., Kay J. H., Stelle K. S.*, Phys. Rev. Lett., **38**, 527 (1977).
52. *van Nieuwenhuizen P., Vermaseren J. A. M.*, Phys. Lett., **65B**, 263 (1976).
53. *Deser S., Kay J. H., Stelle K. S.*, Phys. Rev. Lett., **38**, 527 (1977).
54. *Grisaru M. T., van Nieuwenhuizen P., Vermaseren J. A. M.*, Phys. Rev. Lett., **37**, 1662 (1976).
55. *van Nieuwenhuizen P., Vermaseren J. A. M.*, Phys. Rev., **D16**, 298 (1977).
56. *Fishler M., van Nieuwenhuizen P.*, в печати.
57. *Grisaru M. T.*, Phys. Lett., **66B**, 75 (1977).
58. *Tomboulis E.*, Phys. Lett., **67B**, 417 (1977).
59. *Deser S., Kay J. H., Stelle K. S.*, Phys. Rev. Lett. (1977).
60. *Ferrara S., Zumino B.*, CERN preprint (1978).
61. *Ferrara S., van Nieuwenhuizen P.*, Ecole Normale Superieur preprint 78/14, 1978.

62. *DeWitt B. S.*, Phys. Rev. Lett., **13**, 114 (1964).
63. *Хриплович И. Б.*, Ядерная физика, 3, 575 (1966).
64. *Salam A.*, *Strathdee J.*, Nuovo Cimento Lett., **4**, 101 (1971).
65. *Isham C. J.*, *Salam A.*, *Strathdee J.*, Phys. Rev., **D3**, 867 (1971).
66. *Isham C. J.*, *Salam A.*, Phys. Rev., **D5**, 2548 (1972).
67. *Salam A.*, Phys. Rev., **130B**, 1287 (1963).
68. *Strathdee J.*, Phys. Rev., **135B**, 1428 (1964).
69. *Salam A.*, *Delbourgo R.*, Phys. Rev., **135B**, 1398 (1964).
70. *Delbourgo R.*, University of Tasmania preprint, 1977.
71. *Delbourgo R.*, *West P.*, J. Phys., **A10**, 1049 (1977).
72. *Deser S.*, In: Proceedings of the Conference on Gauge Theories and Modern Field Theory, eds. R. Arnowitt, Nath P., MIT Press, Cambridge Massachusetts, 1975.
73. *Weinberg S.*, In: Proceedings of the XVII International Conference on Hygh Energy Physics ed. J. R. Smith, Rutherford Laboratory, Chilton, Didcot, Oxfordshire, III—59, 1974.
74. *Stelle K. S.*, Phys. Rev., **D16**, 953 (1977).
75. *Salam A.*, *Strathdee J.*, Trieste preprint, 1978.
76. *Källen G.*, Helv. Phys. Acta, **25**, 417 (1952).
77. *Lehmann H.*, Nuovo Cimento, **11**, 342 (1954).
78. *Stanley H. E.*, Phys. Rev., **176**, 718 (1968).
79. *Brézin E.*, *Wallace D. J.*, Phys. Rev., **B7**, 1967 (1973).
80. *Wilson K. G.*, Phys. Rev., **D7**, 2911 (1973).
81. *Dolan L.*, *Jackiw R.*, Phys. Rev., **D9**, 3320 (1974).
82. *Abe R.*, Progr. Theor. Phys., **48**, 1414 (1972).
83. *Parisi G.*, *Peliti L.*, Phys. Lett., **41A**, 331 (1972).
84. *Suzuki M.*, Phys. Lett., **54A**, 5 (1972).
85. *Ferrel R. A.*, *Scalapino D. J.*, Phys. Rev. Lett., **29**, 413 (1972).
86. *Ma S.*, Phys. Rev. Lett., **29**, 1361 (1972).
87. *Coleman S.*, *Jackiw R.*, *Politzer H. D.*, Phys. Rev., **D10**, 2491 (1974).
88. *Gross D. J.*, *Neveu A.*, Phys. Rev., **D10**, 3235 (1974).
89. *Abbott L. F.*, *Kang J. S.*, *Schnitze H. J.*, Phys. Rev., **D13**, 2212 (1976).
90. *Parisi G.*, Nucl. Phys., **B100**, 368 (1975).
91. *Tomboulis E.*, Princeton University preprint, 1977.
92. *Lee T. D.*, *Wick G. C.*, Nucl. Phys., **B9**, 209 (1969); Nucl. Phys., **B10**, 1 (1969); Phys. Rev., **D2**, 1033 (1970).
93. *Weinberg S.*, Phys. Lett., **9**, 357 (1964); Phys. Rev., **B135**, 1049 (1964); Phys. Rev., **B138**, 988 (1965); In: Lectures on Particles and Field Theory, eds, S. Deser and K. Ford, p. 988, Prentice-Hall, New Jersey, 1965.
94. *Boulware D.*, *Deser S.*, Ann. Phys., **89**, 173 (1975).
95. *Weinberg S.*, Gravitation and Cosmology—Principles and Applications of the General Theory of Relativity, Section 7.1, Wiley, New York, 1972. [Имеется перевод: *Вейнберг С.* Гравитация и космология.— М.: Мир, 1975.]
96. *Weinberg S.*, In: Proceedings of the XIV International Conference on High Energy Physics, CERN, Geneva, 1968, p. 253.
97. *Weinberg S.*, In: Lectures on Elementary Particles and Quantum Field Theory—1970 Brandeis Summer Institute in Theoretical Physics, eds. S. Deser., M. Grisaru and H. Pendleton, MIT Press, Cambridge, Massachusetts, 1970.
98. *Lee B. W.*, Chiral Dynamics, Gordon and Breach, New York, 1972.
99. *Weinberg S.*, Phys. Rev. Lett., **18**, 507 (1967).
100. *Schwinger J.*, Phys. Lett., **24B**, 473 (1967).
101. *Weinberg S.*, Phys. Rev., **166**, 1568 (1968).
102. *Coleman S.*, *Wess J.*, *Zumino B.*, Phys. Rev., **177**, 2239 (1968).
103. *Callan C.*, *Coleman S.*, *Wess J.*, *Zumino B.*, Phys. Rev., **177**, 2247 (1968).
104. *Adler S. L.*, *Lieberman J.*, *Ng Y. J.*, *Tsao H.*, Phys. Rev., **D14**, 359 (1976).
105. *Adler S. L.*, Phys. Rev., **D14**, 379 (1976).
106. *Phillips P. R.*, Phys. Rev., **146**, 966 (1966).
107. *Сахаров А. Д.*, ДАН СССР, **177**, 70 (1967).

108. *Ohanian H. C.*, Phys. Rev., **184**, 1305 (1969).
109. *Dürr H. P.*, Gen. Relativ. Grav., **4**, 29 (1973).
110. *Alkatz D.*, Stony Brook preprint ITP-SB-77-59, 1977.
111. *Teragawa H.*, *Chikashige Y.*, *Akama K.*, *Matsuki T.*, Phys. Rev., **D15**, 1181 (1977).
112. *Matsuki T.*, Progr. Theor. Phys., **59**, 235 (1978).
113. *Akama K.*, *Chikashige Y.*, *Matsuki T.*, Progr. Theor. Phys., **59**, 653 (1978).
114. *Akama K.*, *Chikashige Y.*, *Matsuki T.*, *Terazawa H.*, INS-Report-304, 1977.
115. *Akama K.*, Saitama preprint, 1978.
116. *Weinberg S.*, Phys. Rev., **D8**, 3497 (1973).
117. *Kinoshita T.*, J. Math. Phys., **3**, 650 (1962).
118. *Lee T. D.*, *Nauenberg M.*, Phys. Rev., **133**, B 1549 (1964).
119. *Sterman G.*, *Weinberg S.*, Phys. Rev. Lett., **39**, 1436 (1977).
120. *DeWitt B. S.*, Phys. Rev., **162**, 1195, 1239 (1967).
121. *DeWitt B. S.*, Phys. Rep., **19**, 295 (1975).
122. *'t Hooft G.*, *Veltman M.*, Ann. Inst. Poincaré, **20**, 69 (1974).
123. *Honerkamp J.*, Nucl. Phys., **B48**, 269 (1972).
124. *Kallosh R.*, Nucl. Phys., **B78**, 293 (1974).
125. *Grisaru M. T.*, *van Nieuwenhuizen P.*, *Wu C. C.*, Phys. Rev., **D12**, 3203 (1975).
126. *Weinberg S.*, invited talk at the Eighth International Conference on General Relativity, Waterloo, Ontario, Canada, 1977 (не опубликовано).
127. *Fradkin E. S.*, *Vilkovsky G. A.*, In: Proceedings of the XVIII Onternational Conference on High Energy Physics, vol. 2, Sec. T28, JINR, Dubna, 1977; Berne preprint, 1976, I. A. S. preprint 778—IPP, 1978.
128. *Politzer H. D.*, Phys. Rev. Lett., **30**, 1346 (1973).
129. *Gross D. J.*, *Wilczek F.*, Phys. Rev. Lett., **30**, 1343 (1973).
130. *Wilson K. G.*, *Fisher M. E.*, Phys. Rev. Lett., **28**, 240 (1972).
131. *Wilson K. G.*, Phys. Rev. Lett., **28**, 548 (1972).
132. *Bardeen W. A.*, *Lee B. W.*, *Shrock R. E.*, Phys. Rev., **D14**, 985 (1976).
133. *Brézin E.*, *Zinn-Justin J.*, *Le Guillou J. C.*, Phys. Rev., **D14**, 2615 (1976).
134. *Polyakov A. M.*, Phys. Rev. Lett., **59B**, 79 (1975).
135. *Мигдал А.*, ЖЭТФ, **69**, 1457 (1975).
136. *Wegner F. J.*, J. Phys., C. Solid State Physics, **7**, 2098 (1974).
137. *'t Hooft G.*, *Veltman M.*, Nucl. Phys., **B48**, 189 (1972).
138. *Bollini C. G.*, *Giambiagi J. J.*, Phys. Lett., **40B**, 566 (1972).
139. *Ashmore J. F.*, Nuovo Cimento Lett., **4**, 289 (1972).
140. *Abers E. S.*, *Lee B. W.*, Phys. Rep., **9**, 1 (1973).
141. *Taylor J. C.*, Gauge Theories of Weak Interactions, Cambridge University Press, 1976.
142. *Nambu Y.*, Phys. Lett., **26B**, 626 (1968).
143. *Boulware D. G.*, *Brown L. S.*, Phys. Rev., **172**, 1628 (1968).
144. *Prokhorov L. V.*, Phys. Rev., **183**, 1515 (1969).
145. *Duff M. J.*, Phys. Rev., **D9**, 1837 (1974).
146. *Boulware D.*, *Deser S.*, Ann. Phys., **89**, 173 (1975).
147. *Applequist T.*, *Carazzone J.*, Phys. Rev., **D11**, 2856 (1975).
148. *Weinberg S.*, Physics Today, **36**, No. 4 (1977).
149. *Gildener E.*, *Weinberg S.*, Phys. Rev., **D13**, 3333 (1976).
150. *'t Hooft G.*, Nucl. Phys., **B61**, 455 (1973); Nucl. Phys., **B82**, 444 (1973).
151. *Gross D. J.*, In: Methods in Field Theory, eds. R. Balian and J. Zinn-Justin, section 4, North Holland, Amsterdam, 1976.
152. *Collins J. C.*, Phys. Rev., **D10**, 1213 (1974); University of Cambridge thesis (не опубликовано).
153. *Brézin E.*, *LeGuillou J. C.*, *Zinn-Justin J.*, *Nickel B. G.*, Phys. Lett., **44A**, 227 (1973).
154. *Brézin E.*, *LeGuillou J. C.*, *Zinn-Justin J.*, Phys. Rev., **D8**, 2418 (1973).
155. *Duncan A.*, Phys. Lett., **66B**, 170 (1977).
156. *Brown L.*, Phys. Rev., **D15**, 1469 (1977).
157. *Capper D. M.*, *Duff M. J.*, Nucl. Phys., **B82**, 147 (1974).

158. *Capper D. M., Duff M. J., Halpern L.*, Phys. Rev., **D10**, 461 (1974).
159. *Tsao H.-S.*, Phys. Lett., **66**, 79 (1977).
160. *Yang C. N., Mils R. L.*, Phys. Rev., **96**, 141 (1954).
161. *Gastmans R., Kallosh R., Truffin C.*, Препринт ФИАН им. П. Н. Лебедева, 1977.
162. *Gibbons G. W., Hawking S. W.*, Phys. Rev., **D15**, 2752 (1977). [Имеется перевод: Черные дыры, сб. НФФ, вып. 9.— М.: Мир, 1978, с. 310.]
163. *Hawking S. W.*, Phys. Lett., **60A**, 81 (1977).
164. *Arnowitt R., Deser S., Misner C. W.*, In: Gravitation: An Introduction to Current Research, ed. L. Witten, Wiley, New York, 1962; *York J. W.*, Phys. Rev. Lett., **28**, 1082 (1972).
165. *Misner C. W., Thorne K. S., Wheeler J. A.*, Gravitation, Chapter 21, Freeman, San Francisco, 1973. [Имеется перевод: *Мизнер Ч., Торн К. С., Уилер Дж. А.* Гравитация.— М.: Мир, 1977.]
166. *Christensen S. M., Duff M. J.*, Brandeis preprint, 1978.
167. *Callan C., Coleman S., Jackiw R.*, Ann. Phys., **59**, 42 (1970).

ИМЕННОЙ УКАЗАТЕЛЬ

ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ

СОДЕРЖАНИЕ

УВАЖАЕМЫЙ ЧИТАТЕЛЬ!

Ваши замечания о содержании книги, ее оформлении, качестве перевода и другие просим присылать по адресу: 129820, Москва, И-110, ГСП, 1-й Рижский пер., д. 2, изд-во «Мир».

ОБЩАЯ ТЕОРИЯ ОТНОСИТЕЛЬНОСТИ

Под редакцией

Стивена Хокинга и Вернера Израэля
Ст. научн. редакторы В. И. Самсонова, Н. Л. Телеснин
Мл. научн. редакторы В. Н. Цлаф, Г. Г. Сорокина, Р. Х. Зацепина
Художник В. Е. Карпов
Художественный редактор Л. М. Кузнецова
Технический редактор Н. И. Борисова
Корректор Т. П. Пашковская

ИБ № 3321

Сдано в набор 25.01.83.
Подписано к печати 29.06.83.
Формат 60Х90[1]/[16].
Бумага типографская № 2.
Гарнитура литературная. Печать высокая.
Бум. л. 14,50. Усл. печ. л. 29,00. Усл. кр. отт. 29,00. Уч.-изд. л. 31,04. Изд. № 2/2052.
Тираж 10000 экз. Заказ № 1230. Цена 2 р. 40 к.

ИЗДАТЕЛЬСТВО «МИР»
Москва, 1-й Рижский пер., 2

Ордена Октябрьской Революции и ордена Трудового Красного Знамени Первая Образцовая типография имени А. А. Жданова Союзполиграфпрома при Государственном комитете СССР по делам издательств, полиграфии и книжной торговли. Москва, М-54, Валовая, 28

ОБЩАЯ ТЕОРИЯ ОТНОСИТЕЛЬНОСТИ